# Tests d'indépendance en analyse multivariée et tests de normalité dans les modèles ARMA

par

**Pierre LAFAYE DE MICHEAUX**

Thèse de doctorat effectuée en cotutelle
au

Département de mathématiques et de statistique

Faculté des arts et des sciences

**Université de Montréal**

ET

Département des sciences mathématiques

Formation Doctorale Biostatistique

École Doctorale Information, Structures, Systèmes

**Université Montpellier II**
**Sciences et Techniques du Languedoc**

Thèse présentée à la Faculté des études supérieures de l'Université de Montréal

en vue de l'obtention du grade de Philosophiæ Doctor (Ph.D.) en statistique

et à

l'Université Montpellier II en vue de l'obtention du grade de Docteur
d'Université en Mathématiques appliquées et applications des mathématiques

décembre 2002

**Université de Montréal**
Faculté des études supérieures
et
**Université Montpellier II**
Laboratoire de Probabilité et Statistiques

Cette thèse de doctorat effectuée en cotutelle et intitulée

# Tests d'indépendance en analyse multivariée et tests de normalité dans les modèles ARMA

a été présentée et soutenue publiquement à l'Université de Montréal par

## Pierre LAFAYE DE MICHEAUX

Elle a été évaluée par un jury composé des personnes suivantes :

Président-rapporteur
*Université de Montréal* — *Cléroux R., Professeur, Département de mathématiques et de statistique*

Directeur de recherche
*Université de Montréal* — *Bilodeau M., Professeur, Département de mathématiques et de statistique*

Directeur de recherche
*Université Montpellier II* — *Ducharme G., Professeur, Département des sciences mathématiques*

Examinateur externe
*University of Toronto* — *Feuerverger A., Professor, Department of Statistics*

Examinateur externe*
*Université de Marne la Vallée* — *Diebolt J., Professeur, Laboratoire d'Analyse et de Mathématiques Appliquées*

Membre du jury
*Université Montpellier II* — *Escoufier Y., Professeur, Département des sciences mathématiques*

*Représentant de
l'examinateur externe — Roch Roy, Professeur, Département des sciences mathématiques, Université de Montréal

Représentant du doyen de
la FES, *Université de Montréal* — *Cléroux R., Professeur, Département de mathématiques et de statistique*

Thèse acceptée le:
*16 décembre 2002*

أَلنَجَابَةُ أَلجَدِيدَةُ هِي أَلمَعرِفَةُ

L’art de la connaissance et l’exercice des vertus,
éternelle noblesse du chercheur.

*À ma femme Dominique Delseny et notre fils Luka.*

*À mon meilleur ami Fabien Baudrier.*

*À mes parents.*

# REMERCIEMENTS

Les premières personnes que je tiens à honorer ici sont mes deux directeurs de recherche Martin Bilodeau et Gilles Ducharme.

Je considère Gilles Ducharme comme mon père spirituel en statistique ; il m'a appris l'essentiel de ce que j'en sais. Depuis son arrivée à Montpellier, période à laquelle j'ai entamé mes études dans ce domaine, j'ai commencé à apprécier son approche innovante de la statistique au travers de ses enseignements beaucoup plus intuitifs que lourdement formels. Il m'a ensuite initié à la recherche lors de mon année passée en DEA. Après avoir goûté à sa grande expérience, à l'originalité de ses sujets de recherche et à la pertinence de son jugement pour en maîtriser les difficultés techniques menant à leur résolution, j'ai replongé avec plaisir pour quelques années supplémentaires ! Il m'a enseigné un métier passionnant. Je lui suis aussi reconnaissant pour son soutien moral dans certaines périodes difficiles, pour les contributions financières et matérielles apportées et pour la chance qu'il m'a donnée de pouvoir travailler avec un chercheur de grande valeur à Montréal.

Martin Bilodeau a pris le relais avec beaucoup d'attention et de professionnalisme. J'ai apprécié sa constante fiabilité, sa précieuse disponibilité et sa discrétion. Je lui suis également reconnaissant pour les connaissances scientifiques qu'il m'a transmises et pour les ressources financières dont j'ai bénéficié. A son contact d'une grande probité intellectuelle, j'ai pu apprécier et m'imprégner de son souci d'efficacité et de concision. Je veux lui témoigner ici toute mon estime.

Je tiens à exprimer ma reconnaissance aux membres du jury qui ont accepté de prendre du temps pour lire mon travail, pour leurs conseils judicieux et leurs remarques pertinentes ainsi que pour s'être déplacés pour assister à ma soutenance, ce rite d'acceptation dans une nouvelle communauté.

Merci aux chercheurs avec qui j'ai pu avoir des discussions scientifiques profitables tant à Montpellier (Benoît Cadre, Michel Cuer, Pierre Jacob, Irène Larramendy-Valverde) qu'à Montréal (Jean-Francois Angers, Anne Bourlioux, Richard Duncan, Marlene Frigon, Martin Goldstein, Michel Grundland, Anatole Joffe, Christian Léger, Jean-Marc Lina, Urs Maag, Éric Marchand, François Perron, Roch Roy), sans oublier ceux plus anonymes avec qui j'ai échangé des idées sur Internet.

Il est clair aussi que le travail de recherche effectué au cours d'un doctorat doit s'appuyer sur des bases techniques solides. Les principaux acteurs de l'excellente formation en statistique que j'ai suivie sont Martin Bilodeau, Yves Lepage et François Perron à Montréal et Denis Allard, Alain Berlinet, Gilles Caraux, Jean-Pierre Daures, Gilles Ducharme, Jean-François Durand, Ali Gannoun, Pierre Jacob, Jean-Dominique Lebreton, Pascal Monestiez, Robert Sabatier, Gilberte Vignau et Jean-Pierre Vila à Montpellier.

J'ai bénéficié pendant ces « quatre » années, dont trois consacrées à la recherche, d'un précieux soutien informatique de la part de Christopher Albert, Nicolas Beauchemin, Miguel Chagnon, Baptiste Chapuisat, Brigitte Charnomordic, Marc Fangeaud, Michel Lamoureux, Nathalie Malo, Pascal Neveu, Philippe Vismara ainsi que de Leslie Lamport, Linus Torvalds et des milliers de bénévoles oeuvrant à la conception et au développement de Latex, Linux et autres logiciels libres d'excellence.

Je veux également souligner l'appui administratif de qualité apporté par Jacques Bélair, Robert Cléroux, Véronique Hussin, Sabin Lessard, Thérèse Ouellet, Danièle Proulx, Jacqueline Reggiori-Lamoureux, Yvan Saint-Aubin, Danièle Tessier et Janet Zaki à l'Université de Montréal et par Pierrette Arnaud, Michel Averous, Yves Escoufier, Daniel Guin, Bernadette Lacan, Marie-Odile Morda, Florence Picone et Véronique Sals-Vettorel à Montpellier.

Tout ce travail de recherche aurait été certainement plus long, pénible et de moins bonne qualité, si je n'avais pas bénéficié de plusieurs financements octroyés généreusement par le Département de mathématiques et de statistique et la Faculté des Études Supérieures de l'Université de Montréal, par l'Université de Montréal, par l'Institut des Sciences Mathématiques de Montréal ainsi que par mes deux directeurs de recherche Martin Bilodeau et Gilles Ducharme. Merci au Gouvernement du Canada et à celui du Québec pour m'avoir donné ma chance.

Toute ma gratitude aux personnes rencontrées à l'Université Montpellier II et dont une bonne partie m'ont offert leur amitié : Gérard Biau pour ses précieux conseils et son amitié depuis l'année de DEA, Sandie Ferrigno pour son soutien au site du DEA, Benoît Frichot pour les intéressantes discussions et son soutien moral, Olivier Gimenez (il y aurait trop à dire ici), Mariem Mint-el-Mouvid pour les bons moments passés dans notre bureau 14 malgré nos difficultés financières, Nicolas Molinari pour ses précieux conseils et les bons moments passé à Sauve. Une pensée à Omar Anza, Élodie Brunel, André Diatta, Bénédicte Fontez, Laurent Gardes et Hassan Mouhanad. Toutes ces personnes ont contribué, à leur degré, à animer l'espace de création du bâtiment 9. Merci aussi à Alain Delcamp, Ali Gannoun et Jérome Saracco pour les bons moments passés au 3ème étage et à Stéphane Girard et Cécile Amblard pour leur accueil très chaleureux à Montréal.

Enfin merci à tous mes nouveaux amis rencontrés à Montréal : Yves Atchade (Bénin), Chafik Bouhaddioui (Maroc), Alain Desgagné (Québec), Alexandre Leblanc (Québec), Ghislain Rocheleau (Québec), Ndeye Rokhaya Gueye (Sénégal) qui ont fait une partie de leur doctorat en même temps que moi ; et aussi Christopher Albert (USA)

et Carole (France), Marie-Soleil Beaudoin (Québec), Pascal Bouchard (Québec), Jean-Francois Boudreau (Québec), Pascal Croteau (Québec), Alexandre Cusson (Québec), Alina Dyachenko (Russie), Alexis Gerbeau (Québec), Mohammed Haddou (Algérie), Hassiba et Djamal Hellal (Algérie), Abdelaziz Khatouri (Maroc), Vincent Lemaire (France), Nathalie Malo (Québec), Hacène Nedjar (Algérie), Philippe Paradis (Québec), Fritz Pierre (Haiti), Alice Savu (Roumanie), Jib et Sarah. Vous avez facilité mon intégration. J'espère garder des liens solides avec la plupart d'entre vous.

# Table des matières

# LISTE DES TABLEAUX

# LISTE DES FIGURES

# SIGLES ET ABRÉVIATIONS

$\xrightarrow{L}$ : Convergence en loi.
$\xrightarrow{fd}$ : Convergence des lois de dimension finie.
$\Rightarrow$ : Convergence faible.
ARMA : Auto-Regressive Moving Average.
*i.i.d.* : Indépendant(e)s et identiquement distribué(e)s.
$\equiv$ : signe d'égalité fonctionnelle.
$\langle \cdot , \cdot \rangle$ : Produit scalaire.
$\Omega$ : Espace des évenements.
$\mathsf{P}$ : Mesure de probabilité sur l'espace des évenements.
$\pi_{\boldsymbol{t}_1,\ldots,\boldsymbol{t}_k}$ : Projection.
$C \equiv C(\mathbb{R}^p, \mathbb{C})$ : Espace des fonctions continues de $\mathbb{R}^p$ dans $\mathbb{C}$.
$\rho$ : Métrique sur $C(\mathbb{R}^p, \mathbb{C})$.
*m.l.e.* : Maximum likelihood estimator.
$^\mathsf{T}$ : Transposée d'un vecteur.
MANOVA : Multivariate analysis of variance.
EDF : Empirical distribution function.
CF : Cubature formula.

# CHAPITRE 1

# Introduction

*Une procédure conſiſtant à déterminer ſi un modèle probabiliſte particulier eſt approprié pour un phénomène aléatoire donné…*

Depuis le tout début de la statistique, nombre de statisticiens ont commencé leur analyse en proposant une distribution pour leurs observations et ont ensuite tenté de vérifier si leur distribution était la bonne. Ainsi, au fil des ans, un grand nombre de telles procédures sont apparues commençant alors à constituer un vaste champ d'études portant le nom de « tests d'ajustement ». Il est important de noter que, si l'on ne disposait pas des outils que sont les tests d'ajustement, il faudrait se baser sur des critères subjectifs pour valider la qualité d'un modèle. Malheureusement, comme l'a si bien souligné Fisher (1925)

*No eye observation of such diagrams, however experienced, is really capable of discriminating whether or not the observations differ from the expectation by more than we would expect from the circumstances of random sampling.*

L'apport des tests d'ajustement est triple. Il permet d'obtenir une description compacte des données en leur attribuant une loi de probabilité. Ensuite, certaines techniques paramétriques puissantes sont valides uniquement sous l'hypothèse de normalité. Enfin, cela permet de mieux comprendre le mécanisme ayant généré les données, en obtenant de l'information sur les raisons ayant conduit à un rejet de l'hypothèse de travail.

Mathématiquement, le problème se présente de la façon suivante. Soit $Y$ un élément aléatoire dont la fonction de répartition $F$ est absolument continue par rapport à la mesure de Lebesgue. On désire tester l'hypothèse

$$H_0 : F(x) \in \{F_0(x, \boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta\} \text{ versus } H_1 : F(x) \notin \{F_0(x, \boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta\},$$

où $\Theta$ est un certain espace paramétrique. Pour ce problème, on peut distinguer deux grandes classes de tests. Les tests « omnibus » concernent les situations où l'on n'a *a priori* aucune indication sur la façon dont la distribution réelle $F$ pourrait s'écarter de l'hypothèse nulle. Ils sont efficaces contre des alternatives non spécifiées et sont généralement basés sur la fonction de répartition expérimentale ou sur la fonction caractéristique expérimentale. On peut citer par exemple les tests de Kolmogorov, d'Anderson-Darling ou de Cramér-von Mises. Les tests « directionnels » quant à eux permettent de prendre en compte certaines informations sur les écarts les plus plausibles à l'hypothèse nulle. Ils sont construits de façon à détecter avec plus de puissance certains types d'orientation que pourrait prendre la distribution de $Y$. Au cours de la première partie de la recherche envisagée, on s'est intéressé à la seconde classe en adaptant la théorie des tests lisses, introduite par Neyman (1937), au contexte particulier de données dépendantes, issues d'une loi non entièrement spécifiée.

En effet, une autre branche importante et originale de la statistique concerne l'analyse des séries chronologiques. Sa caractéristique essentielle réside en une dépendance des phénomènes étudiés vis-à-vis du temps, concept essentiel tant au niveau scientifique que philosophique. Il est de fait peu de disciplines qui ne soient confrontées à l'étude de variables évoluant dans le temps et qu'on désire décrire, expliquer, contrôler ou prévoir. Cette discipline puise ses origines dans le Moyen Âge comme en témoigne ce diagramme temporel (représentant l'inclinaison des orbites de planètes en fonction du temps) considéré comme l'un des plus anciens du monde occidental.

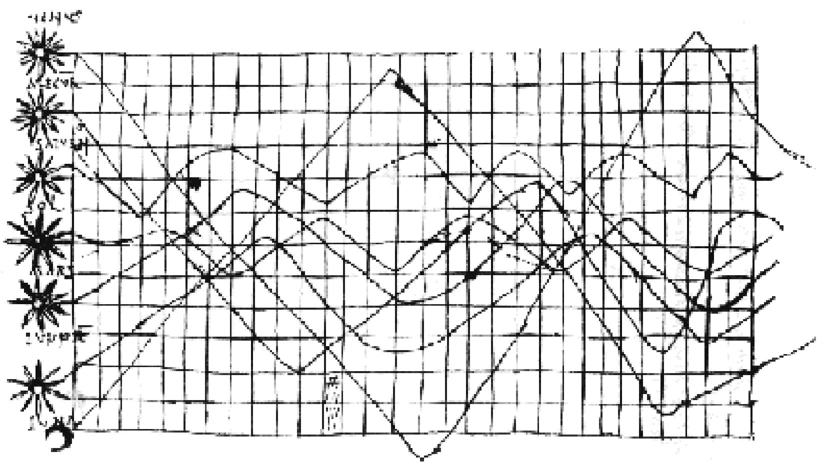

FIG. 1.1. 10th Century time line (Funkhauser, 1936, pp 260-262)

Une avancée importante dans l'étude des séries temporelles a été de supposer que la série chronologique observée est engendrée par un processus stochastique $\{Y_t, t \in \mathbb{Z}\}$. Une condition souvent imposée sur ce processus générateur est qu'il soit stationnaire du second ordre. Un processus digne d'intérêt satisfaisant à ces conditions est le processus autorégressif à moyenne mobile (ARMA). Ce processus est très utilisé du fait de sa simplicité. Son introduction nécessite quelques définitions préalables.

**Définition 1.** *Stationnarité faible*
*Un processus $\{Y_t; t \in \mathbb{Z}\}$ est dit stationnaire au second ordre, ou stationnaire au sens faible, ou stationnaire d'ordre deux si les trois conditions suivantes sont satisfaites :*

– $\forall t \in \mathbb{Z}, \quad E(Y_t^2) < \infty,$

– $\forall t \in \mathbb{Z}, \quad E(Y_t) = m$*, indépendant de t,*
– $\forall (t,h) \in \mathbb{Z}^2, \quad cov(Y_t, Y_{t+h}) = E[(Y_{t+h} - m)(Y_t - m)] = \gamma(h)$*, indépendant de t.*

Nous avons aussi besoin de la notion de bruit blanc (ou white noise).

**Définition 2.** *Bruit blanc*
*Un processus* $\{\epsilon_t; t \in \mathbb{Z}\}$ *est un bruit blanc s'il satisfait aux deux conditions suivantes* $\forall t \in \mathbb{Z}$ *:*
– $E(\epsilon_t) = 0$,
– $\gamma(h) = E[\epsilon_t \epsilon_{t-h}] = \begin{cases} \sigma^2, \ h = 0, \\ 0, \ h \neq 0. \end{cases}$

En s'appuyant sur ces deux définitions, on peut introduire le modèle ARMA.

**Définition 3.** *Modèle ARMA(p,q)*
*On appelle processus autorégressif à moyenne mobile d'ordre (p,q) un processus stationnaire* $\{Y_t; t \in \mathbb{Z}\}$ *vérifiant une relation du type*

$$Y_t - \sum_{i=1}^{p} \varphi_i Y_{t-i} = \sum_{i=1}^{q} \theta_i \epsilon_{t-i} + \epsilon_t \ , \forall t \in \mathbb{Z}$$

*où les* $\varphi_i$ *et les* $\theta_i$ *sont des réels et où l'erreur* $(\epsilon_t; t \in \mathbb{Z})$ *est un bruit blanc de variance* $\sigma^2$.

Dans la démarche de modélisation d'une série temporelle, il est courant de suivre la procédure indiquée par Box et Jenkins (1976). Cette procédure se déroule en cinq étapes : « Stationnarisation », « Désaisonnalisation », « Identification », « Validation et Tests », « Prévisions ».

L'attention dans la première partie de cette étude s'est portée sur la phase de « Validation et Tests », et plus particulièrement sur la construction d'un test lisse de normalité pour les erreurs d'un modèle ARMA($p$,$q$) univarié complètement identifié, de moyenne connue.

Pour cela, la statistique du *score de Rao* est un outil classique pour la construction de tests d'hypothèses de la forme $H_0 : \boldsymbol{\eta} = \boldsymbol{\eta}_0$. Le test qui en découle est basé sur le principe général que le vecteur gradient du modèle non restreint (par $H_0$), évalué en l'estimateur restreint, suit asymptotiquement une loi normale de moyenne 0, si $H_0$ est vraie. Si l'hypothèse alternative est décrite par une certaine famille exponentielle de dimension $K$, le test du score résultant est aussi appelé test lisse (*smooth test*) ou test lisse de Neyman (1937) d'ordre $K$. L'idée du test lisse est en fait d'emboîter la fonction de densité de l'hypothèse nulle dans une famille paramétrique plus générale. Cette famille doit être choisie pour détecter les alternatives les plus probables si l'hypothèse nulle est fausse. De cette façon, l'hypothèse nulle devient une hypothèse paramétrique où l'on cherche à tester la nullité de paramètres de la densité alternative. Cette approche fournit non seulement une statistique de test simple, mais aussi une bonne puissance pour une vaste famille d'alternatives. Considérons un échantillon $\epsilon_1, \ldots, \epsilon_n$ de variables aléatoires continues *i.i.d.* ayant pour fonction de répartition $F$ et pour lequel on souhaite tester l'hypothèse $H_0 : F = F_0$. Il est possible de transformer cet échantillon en un échantillon $U_1 = 2F_0(\epsilon_1) - 1, \ldots, U_n = 2F_0(\epsilon_n) - 1$ de variables aléatoires *i.i.d.* de densité $g$ qui, sous $H_0$, sont de loi uniforme sur l'intervalle [-1,1]. Utilisant cette propriété, Neyman (1937) propose de considérer l'hypothèse

$$H_0 : g(y) = \begin{cases} 1/2 & \text{si } y \in [-1, 1], \\ 0 & \text{sinon} \end{cases} \tag{1}$$

et de choisir pour cette loi uniforme la famille alternative d'imbrication de fonctions de densité $c. \exp[P(y)]$ où $P(y)$ représente un polynôme et $c$ est la constante de normalisation. Il appelle de telles densités des alternatives « lisses » car elles sont représentées graphiquement par des courbes lisses coupant la densite sous $H_0$ un petit nombre de fois. Le fait de pouvoir se restreindre, par des considérations physiques ou en faisant appel à son intuition, à ce type d'alternatives permet d'obtenir un test plus sensible. Ensuite, il suggère de choisir le polynôme $P(y)$ comme étant une combinaison linéaire des éléments d'un des systèmes de polynômes orthonormaux $\pi_0(y), \pi_1(y), \ldots$ sur l'intervalle [-1,1]. Rappelons que, dans un tel système, le polynôme $\pi_i(y)$ est de degré $i$, et que pour tout $i$ et $j = 0, 1, 2, \ldots$, on a

$$1/2 \int_{-1}^{1} \pi_i(y)\pi_j(y)dy = \delta_{ij}$$

où

$$\delta_{ij} = \begin{cases} 0 & \text{si } i \neq j, \\ 1 & \text{si } i = j. \end{cases}$$

Ces systèmes de polynômes orthonormaux peuvent être construits de plusieurs façons (Marsden, 1974, p. 347). Neyman (1937), pour sa part, utilise le système de polynômes de Legendre. La famille d'imbrication s'écrit alors

$$g_K(y, \boldsymbol{\eta}) = \begin{cases} C(\boldsymbol{\eta}) \exp\left(\sum_{i=1}^{K} \eta_i \pi_i(y)\right) & \text{si } y \in [-1, 1], \\ 0 & \text{sinon.} \end{cases} \tag{2}$$

où $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_k)^t \in \mathbb{R}^K$, $C(\boldsymbol{\eta})$ est la constante de normalisation, $K$ est appelé l'ordre de l'alternative et les $\pi_i(\cdot), i = 1, \ldots, K$ sont les $K$ premiers polynômes de Legendre. Nous dénotons par $g_0(\cdot)$ la fonction de densité sous l'hypothèse nulle, c'est-à-dire la distribution uniforme sur l'intervalle [-1,1]. Ainsi, éprouver l'hypothèse nulle (1) contre l'hypothèse voulant que la fonction de densité appartienne à la famille d'imbrication $H_1 : g(\cdot) \in \{g_k(\cdot, \boldsymbol{\eta}) : \boldsymbol{\eta} \in \mathbb{R}^K\}$, donnée en (2), est donc équivalent à éprouver l'hypothèse

$$H_0 : \boldsymbol{\eta} = 0 \text{ versus } H_1 : \boldsymbol{\eta} \neq 0. \tag{3}$$

Il suffit alors de construire le test du $score$ pour l'hypothèse (3). On définit le vecteur du $score$ par

$$\boldsymbol{a}_{n\boldsymbol{\eta}} = \left( \frac{1}{n} \sum_{i=1}^{n} \frac{\partial Log g_K(Y_i, \boldsymbol{\eta})}{\partial \eta_1}, \ldots, \frac{1}{n} \sum_{i=1}^{n} \frac{\partial Log g_K(Y_i, \boldsymbol{\eta})}{\partial \eta_K} \right)^{\mathsf{T}}.$$

Alors la statistique du score de Rao (1947) est

$$R_K = n\boldsymbol{a}_{n\boldsymbol{\eta}_0}^{\mathsf{T}} I_{\boldsymbol{\eta}_0}^{-1} \boldsymbol{a}_{n\boldsymbol{\eta}_0}$$

où

$$I_{\boldsymbol{\eta}} = \left[ E_{\boldsymbol{\eta}} \left\{ \frac{\partial Logg_K(Y, \boldsymbol{\eta})}{\partial \eta_i} . \frac{\partial Logg_K(Y, \boldsymbol{\eta})}{\partial \eta_j} \right\} \right]_{K \times K}$$

est la matrice d'information de Fisher de $\boldsymbol{\eta}$. On peut ensuite montrer que, sous $H_0$, $R_K \xrightarrow{L} \chi^2_K$, loi du khi-deux à $K$ degrés de liberté.

Cette présentation du test de Neyman (1937) repose sur l'hypothèse que $F_0$ est entièrement spécifiée. Une généralisation au cas où un ou plusieurs paramètres seraient inconnus a été faite par Kopecky et Pierce (1979) et Rayner et Best (1988).

**Remarque 1.** *Une justification théorique que l'on peut offrir pour le choix de la famille d'imbrication précédente repose sur la théorie des espaces de Hilbert (Royden, 1968, chap.10 sec.8). En effet, on peut montrer que toute fonction $h$ dans l'espace de Hilbert des fonctions de carré intégrable par rapport à $g_0(y)$ peut s'écrire*

$$h(y) = \sum_{i=0}^{\infty} \eta_i \pi_i(y).$$

*Par conséquent, la vraie distribution de Y peut s'écrire*

$$f(y) = \begin{cases} c.exp\left(\sum\limits_{i=0}^{\infty} \eta_i \pi_i(y)\right) & \textit{si } y \in [-1, 1], \\ 0 & \textit{sinon,} \end{cases}$$

*et la famille d'imbrication proposée est une approximation de la vraie densité, d'autant meilleure que $K$ est grand.*

Il est possible d'adapter ce qui précède au cas de données issues d'un processus de type ARMA, c'est l'objet du Chapitre 2.

Le Chapitre 3 est consacré à un problème connexe. On pourrait en effet supposer, dans le même ordre d'idée de la problématique précédente, que les erreurs $\epsilon_t$ du modèle ARMA sont Gaussiennes et se demander si ces erreurs sont engendrées par un processus bruit blanc ou, ce qui est équivalent dans ce cas de figure, s'il y a indépendance sérielle entre les $\epsilon_t$. Dans l'optique d'une future généralisation des résultats du Chapitre 2 au cas d'un ARMA multivarié, il semble intéressant de construire un test d'indépendance sérielle pour des vecteurs aléatoires. De plus, afin de mieux appréhender la complexité du problème, il s'avère avantageux d'échelonner la réponse à cette question en commençant par bâtir un test d'indépendance dans le cas non sériel.

La construction de ce test non paramétrique s'appuie sur une caractérisation de l'indépendance introduite par Ghoudi et al. (2001) et le résultat obtenu est une statistique de type Cramér-von Mises d'un certain processus empirique. Ghoudi et al. (2001) ont défini leur processus en utilisant la fonction de répartition empirique. Ici, le processus empirique est basé sur la fonction caractéristique empirique multivariée. Pour être plus formel, considérons le vecteur aléatoire $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}^{(1)}, \ldots, \boldsymbol{\epsilon}^{(p)})$, constitué de $p$ sous-vecteurs de dimension $q$ et le vecteur $\boldsymbol{t} = (\boldsymbol{t}^{(1)}, \ldots, \boldsymbol{t}^{(p)})$ partitionné de la même façon. En outre, pour tout $A \subset \{1, \ldots, p\}$ de cardinalité supérieure à 1, introduisons la fonction $\mu_A$ définie par

$$\mu_A(\boldsymbol{t}) = \sum_{B \subset A} (-1)^{|A \backslash B|} C_p(\boldsymbol{t}^B) \prod_{j \in A \backslash B} C^{(j)}(\boldsymbol{t}^{(j)})$$

avec

$$(\boldsymbol{t}^B)^{(i)} = \begin{cases} \boldsymbol{t}^{(i)} & i \in B, \\ \mathbf{0} & i \in I_p \backslash B. \end{cases}$$

L'objet $C_p$ est la fonction caractéristique conjointe de $\boldsymbol{\epsilon}$ et les $C^{(j)}$ sont les fonctions caractéristiques des marginales $\boldsymbol{\epsilon}^{(j)}$. Il est alors possible de montrer que $\boldsymbol{\epsilon}^{(1)}, \ldots, \boldsymbol{\epsilon}^{(p)}$ sont indépendants si et seulement si $\mu_A \equiv 0$. Disposant d'un échantillon $\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_n$, cela amène assez naturellement à définir le processus $R_{n,A}$, sous l'hypothèse de multinormalité des $\boldsymbol{\epsilon}_i^{(j)}$, par

$$R_{n,A}(\boldsymbol{t}) = \sqrt{n} \sum_{B \subset A} (-1)^{|A \backslash B|} \phi_{n,p}(\boldsymbol{t}^B) \prod_{i \in A \backslash B} \phi(\boldsymbol{t}^{(i)}) \tag{4}$$

où

$$\phi_{n,p}(\boldsymbol{t}) = \frac{1}{n} \sum_{j=1}^{n} \exp(\mathrm{i}\langle \boldsymbol{t}, \boldsymbol{\epsilon}_j \rangle)$$

est la fonction caractéristique empirique de l'échantillon et $\phi$ est la fonction caractéristique d'une loi normale $N_q(\mathbf{0}, \boldsymbol{I})$. La statistique de Cramér-von Mises et le test qui en découle seront construits à partir de ce processus. Introduisons maintenant quelques définitions et notations utiles pour la suite portant sur la convergence faible d'éléments aléatoires dans un espace métrique $S$.

**Définition 4.** *Convergence en loi*
*La loi de $X$ est par définition la mesure de probabilité $P = \mathsf{P}X^{-1}$ sur $(S, \mathcal{S})$ :*

$$P(A) = \mathsf{P}(X^{-1}(A)) = \mathsf{P}\{\omega \in \Omega; X(\omega) \in A\} = \mathsf{P}\{X \in A\},\ A \in \mathcal{S}.$$

*On dit qu'une suite d'éléments aléatoires $\{X_n\}$ converge en loi vers un élément aléatoire $X$ sur $S$, et on écrit $X_n \xrightarrow{L} X$, si les lois $P_n$ des $X_n$ convergent faiblement vers la loi $P$ de $X$ et on note $P_n \Rightarrow P$.*

Le théorème qui suit (Billingsley, 1968, p. 30) est très utile dans la pratique puisqu'il permet de prouver la convergence faible des mesures induites sur $\mathbb{R}$ par différentes fonctions réelles $h$ à partir de la convergence faible dans des espaces métriques généraux.

**Théorème 1.** *Soit $h : S \longrightarrow S'$ mesurable, et soit $D_h$ l'ensemble des points de discontinuité de $h$.*
*Si $P_n \Rightarrow P$ et $P(D_h) = 0$, alors $P_n h^{-1} \Rightarrow P h^{-1}$.*

On va maintenant donner une caractérisation pratique de la convergence faible mais pour cela on a préalablement besoin d'introduire les deux définitions suivantes.

**Définition 5.** *Compacitude relative*
*Soit $\Pi$ une famille de mesures de probabilité sur $(S, \mathcal{S})$. On dit que $\Pi$ est relativement compacte si toute suite d'éléments de $\Pi$ contient une sous-suite faiblement convergente.*

**Définition 6.** *Lois de dimension finie*
*Soit $S = C(E, F)$ l'ensemble des fonctions continues de $E$ dans $F$ et $\mathcal{S}$ sa tribu borélienne. Soit $\{P_n\}$ une famille de mesures de probabilité sur $\mathcal{S}$. Les lois de dimension finie des $P_n$ sont les mesures $P_n\pi^{-1}_{\boldsymbol{t_1},\ldots,\boldsymbol{t_k}}$, $\forall k = 1, 2, \ldots, \forall \boldsymbol{t}_1, \ldots, \boldsymbol{t}_k \in E$ où*

$$\begin{array}{rcl} \pi_{\boldsymbol{t_1},\ldots,\boldsymbol{t_k}} : C(E,F) & \longrightarrow & F^k \\ x & \mapsto & (x(\boldsymbol{t_1}), \ldots, x(\boldsymbol{t_k})). \end{array}$$

*On dit que les lois de dimension finie des $P_n$ convergent faiblement vers celles de $P$ si $P_n\pi^{-1}_{\boldsymbol{t_1},\ldots,\boldsymbol{t_k}} \Rightarrow P\pi^{-1}_{\boldsymbol{t_1},\ldots,\boldsymbol{t_k}}$. Lorsque $P_n$ et $P$ sont les lois d'éléments aléatoires $X_n$ et $X$ on écrit aussi $X_n \xrightarrow{fd} X$.*

Il est facile de montrer que $P_n \Rightarrow P$ si et seulement si les lois de dimension finie des $P_n$ convergent faiblement vers celles de $P$ et $\{P_n\}$ est relativement compacte. Une façon simple de démontrer la compacitude relative d'une famille de mesures est d'utiliser la notion de tension de cette famille.

**Définition 7.** *Tension d'une famille*
*Une famille $\Pi$ de mesures de probabilité sur un espace métrique général $S$ est dite tendue si*

$$\forall \epsilon > 0,\ \exists K \textit{ compact tel que } P(K) > 1 - \epsilon,\ \forall P \in \Pi.$$

*Une famille d'éléments aléatoires $\{X_n\}$ est dite tendue si la famille des lois des $X_n$ est tendue.*

Le résultat liant les deux concepts précédents peut être résumé ainsi.

**Théorème 2.** *Théorème de Prohorov*
*Si $\Pi$ est tendue, elle est relativement compacte. Si S est séparable et complet et si $\Pi$ est relativement compact, elle est tendue.*

Dans cette optique, un théorème très utilisé, par exemple pour démontrer la convergence en loi de la statistique « classique » de Cramér-von Mises est le suivant.

**Théorème 3.** *(Théorème 8.1, Billingsley (1968))*
*Soient $P_n$, $P$ des mesures de probabilité sur $(C[0, 1], \mathcal{C})$. Si les lois de dimension finie des $P_n$ convergent faiblement vers celles de $P$, et si $\{P_n\}$ est tendue, alors $P_n \Rightarrow P$.*

Dans le cas qui nous préoccupe ici, la fonctionnelle de Cramér-von Mises, basée sur la fonction caractéristique empirique, n'est même pas définie sur $C(\mathbb{R}^{pq}, \mathbb{C})$ et on ne pourra pas utiliser le Théorème 1. Pour résoudre ce problème, on a généralisé le Théorème 3.3 de Kellermeier (1980). Par ailleurs, dans le contexte particulier d'un processus basé sur la fonction caractéristique empirique, qui prend ses valeurs dans l'ensemble $\mathbb{C}$ des nombres complexes et dont l'ensemble des indices est $\mathbb{R}^n$, le problème est bien plus compliqué que dans le cas « classique ». Pour cela, il est utile d'introduire les éléments suivants.

Notons $C(\mathbb{R}^{pq}, \mathbb{C})$ l'ensemble des fonctions continues de $\mathbb{R}^{pq}$ dans $\mathbb{C}$. Schématiquement, la correspondance du processus $R_{n,A}$ de (4) peut se traduire ainsi.

$$\begin{aligned}R_{n,A} : (\Omega, \mathcal{B}, \mathsf{P}) &\longrightarrow C(\mathbb{R}^{pq}, \mathbb{C})\\ \omega &\mapsto R_{n,A}(\cdot, \omega) : \quad \mathbb{R}^{pq} \longrightarrow \mathbb{C}\\ &\qquad\qquad\qquad \boldsymbol{t} \mapsto R_{n,A}(\boldsymbol{t}, \omega) \overset{N}{=} R_{n,A}(\boldsymbol{t}).\end{aligned}$$

On définit une métrique $\rho$ sur l'espace de Fréchet séparable $C(\mathbb{R}^{pq}, \mathbb{C})$ par

$$\forall x, y \in C(\mathbb{R}^{pq}, \mathbb{C}),\ \rho(x, y) = \sum_{j=1}^{\infty} 2^{-j} \frac{\rho_j(x, y)}{1 + \rho_j(x, y)}$$

avec $\rho_j(x, y) = \sup_{\|\boldsymbol{t}\| \le j} |x(\boldsymbol{t}) - y(\boldsymbol{t})|$, où $|\cdot|$ désigne le module sur $\mathbb{C}$. Notons que $\rho_j$ est bien définie car une fonction continue sur un compact est bornée.

**Remarque 2.** *Un espace de Fréchet est un espace métrique linéaire complet.*

On note $\mathcal{S}$ la tribu engendrée par les ouverts de $C \equiv C(\mathbb{R}^{pq}, \mathbb{C})$ pour la métrique $\rho$, $\mathbb{R}^{pq}_j$ la boule fermée de centre $\mathbf{0}$ et de rayon $j$ dans $\mathbb{R}^{pq}$, $C_j \equiv C(\mathbb{R}^{pq}_j, \mathbb{C})$ et $\mathcal{S}_j$ la tribu borélienne sur $C_j$ pour la métrique $\rho_j$.
Notons aussi

$$\begin{aligned}r_j : C(\mathbb{R}^{pq}, \mathbb{C}) &\longrightarrow C_j\\ x &\mapsto r_j(x) : \quad \mathbb{R}^{pq}_j \longrightarrow \mathbb{C}\\ &\qquad\qquad \boldsymbol{t} \mapsto x(\boldsymbol{t}) = r_j(x(\boldsymbol{t}))\end{aligned}$$

la restriction d'un élément $x$ de $C(\mathbb{R}^{pq}, \mathbb{C})$ à $C_j$. Puisque $\mathbb{R}^{pq}$ est localement compact, séparable et Hausdorff on a, d'après la Proposition 14.6 de (Kallenberg, 1997, p. 260),

$$R_{n,A} \xrightarrow{L} R_A \text{ si et seulement si } r_j(R_{n,A}) \xrightarrow{L} r_j(R_A),\ \forall j \ge 1.$$

On peut donc restreindre l'étude du processus $R_{n,A}$ aux sous-espaces compacts de $C$. Maintenant, d'après le Lemme 14.2 (Kallenberg, 1997, p. 256), on a
$r_j(R_{n,A}) \xrightarrow{L} r_j(R_A)$ si et seulement si $r_j(R_{n,A}) \xrightarrow{fd} r_j(R_A)$ et $\{r_j(R_{n,A})\}_n$ est une famille relativement compacte. De plus, d'après le Théorème 14.3 (Kallenberg, 1997, p. 257) et puisque $C_j$ est séparable et complet, on a l'équivalence entre la compacitude relative et la tension de $\{r_j(R_{n,A})\}_n$. Pour prouver la convergence faible de $\{R_{n,A}\}_n$ vers l'élément aléatoire $R_A$ il suffit donc de montrer que les lois de dimension finie des $r_j(R_{n,A})$ convergent faiblement vers celles des $r_j(R_A)$ et que $\forall j \ge 1$, $\{r_j(R_{n,A})\}_n$ est tendue.

Tous ces résultats sont largement exploités et détaillés dans le Chapitre 3 pour construire un test semi-paramétrique d'indépendance, pour des marginales multinormales, qui peut être utile par exemple dans l'étude des données familiales. Ils sont ensuite généralisés au cas sériel.

Pour conclure, il apparait donc que notre problématique repose sur des questions ancrées dans les prémisses de la statistique tandis que notre recherche s'appuie sur des outils et techniques récentes et innovatrices empruntées à l'analyse multivariée (voir Bilodeau et Brenner (1999)), à la théorie des processus stochastiques (voir Billingsley (1968)) ainsi qu'aux méthodes asymptotiques (voir Ferguson (1996)).

## BIBLIOGRAPHIE


[1] Billingsley, P., 1968. Convergence of probability measures. John Wiley & Sons Inc., New York.

[2] Bilodeau, M., Brenner, D., 1999. Theory of multivariate statistics. Springer Texts in Statistics. Springer-Verlag, New York.

[3] Box, G. E. P., Jenkins, G. M., 1976. Time series analysis : forecasting and control, revised Edition. Holden-Day, San Francisco, Calif., holden-Day Series in Time Series Analysis.

[4] Ferguson, T. S., 1996. A course in large sample theory. Chapman & Hall, London.

[5] Fisher, R. A., 1925. Statistical methods for research workers. VIII + 239 p. with 6 tables. Edinburgh and London, Oliver and Boyd.

[6] Funkhauser, H. G., 1936. A Note on a Tenth Century Graph. Osiris, Vol.1.

[7] Ghoudi, K., Kulperger, R. J., Rémillard, B., 2001. A nonparametric test of serial independence for time series and residuals. J. Multivariate Anal. 79, 191–218.

[8] Kallenberg, O., 1997. Foundations of modern probability. Springer-Verlag, New York.

[9] Kellermeier, J., 1980. The empirical characteristic function and large sample hypothesis testing. J. Multivariate Anal. 10 (1), 78–87.

[10] Kopecky, K. J., Pierce, D. A., 1979. Efficiency of smooth goodness-of-fit tests. J. Amer. Statist. Assoc. 74, 393–397.

[11] Marsden, J., 1974. Elementary classical analysis. New York : W.H. Freeman and Company.

[12] Neyman, J., 1937. Smooth test for goodness of fit. Skand. Aktuar. 20, 149–199.

[13] Rao, C., 1947. Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. In : Proceedings of the Cambridge Philosophical Society. Vol. 44. pp. 50–7.

[14] Rayner, J., Best, D., 1988. Smooth tests of goodness of fit for regular distributions. Comm. Statist. Theory Methods 17 (10), 3235–67.

[15] Royden, H., 1968. Real analysis. New York : MacMillan.

# CHAPITRE 2

# Goodness-of-fit tests of normality for the innovations in ARMA models

Cet article a été accepté pour publication dans la revue *Journal of Time Series Analysis*.

Comme la coutume dans cette discipline le veut, l'ordre alphabétique des auteurs a été respecté.

Voici la liste des contributions principales de Pierre Lafaye de Micheaux à cet article :

- Recherche bibliographique ayant permis notamment la rédaction de la section 1.
- Démonstration des résultats de la section 2.
- Conception, écriture et validation des programmes Fortran77 pour obtenir la table 3.1.
- Conception, écriture et validation des programmes Fortran77 pour obtenir les résultats de la section 4.
- Conception, écriture et validation de l'exemple.
- Participation à la rédaction.

# Goodness-of-fit tests of normality for the innovations in ARMA models

(abbreviated title: Testing the residuals in ARMA)

Gilles R. Ducharme and Pierre Lafaye de Micheaux

*Laboratoire de probabilités et statistique, cc051*
*Université Montpellier II*
*Place Eugène Bataillon*
*34095, Montpellier, Cedex 5*
*France*

**Abstract**
In this paper, we propose a goodness-of-fit test of normality for the innovations of an ARMA($p$, $q$) model with known mean or trend. This test is based on the data driven smooth test approach and is simple to perform. An extensive simulation study is conducted to study the behavior of the test for moderate sample sizes. It is found that our approach is generally more powerful than existing tests while holding its level throughout most of the parameter space and thus, can be recommended. This agrees with theoretical results showing the superiority of the data driven smooth test approach in related contexts.

*Key words:* ARMA process, Gaussian white noise, Goodness-of-fit test, Normality of residuals, Smooth test.

## 1. Introduction

Let ($Y_t$, $t \in \mathbb{Z}$) be a stationary process. In this paper, we consider the case where $E(Y_t)$ is known or has been estimated using information outside of the data set. Thus, without loss of generality, we set $E(Y_t) = 0$. Consider the framework where ($Y_t$, $t \in \mathbb{Z}$) obeys the causal and invertible finite order ARMA($p$, $q$) model

$$Y_t - \boldsymbol{\varphi}^{\mathsf{T}} \boldsymbol{Y}_{t-1}^{(p)} = \boldsymbol{\theta}^{\mathsf{T}} \boldsymbol{\epsilon}_{t-1}^{(q)} + \epsilon_t \tag{1}$$

where ($\epsilon_t$, $t \in \mathbb{Z}$) is an innovation process of random variables with mean 0 and autocovariance $E(\epsilon_t \epsilon_{t+h}) = \sigma^2 < \infty$ (unknown) if $h = 0$ and 0 otherwise and where

$$\boldsymbol{\varphi} = \begin{bmatrix} \varphi_1 \\ \vdots \\ \varphi_p \end{bmatrix}, \ \boldsymbol{\theta} = \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_q \end{bmatrix}, \ \boldsymbol{Y}_{t-1}^{(p)} = \begin{bmatrix} Y_{t-1} \\ \vdots \\ Y_{t-p} \end{bmatrix}, \ \boldsymbol{\epsilon}_{t-1}^{(q)} = \begin{bmatrix} \epsilon_{t-1} \\ \vdots \\ \epsilon_{t-q} \end{bmatrix}.$$

A sample $\{Y_1, ..., Y_T\}$ is observed and model (1) is fitted by standard methods, for example the unconditional Gaussian maximum likelihood approach (see Brockwell and Davis (1991), p. 256-257), yielding the estimator $\hat{\boldsymbol{\beta}} = (\hat{\boldsymbol{\varphi}}^{\mathsf{T}}, \hat{\boldsymbol{\theta}}^{\mathsf{T}}, \hat{\sigma})^{\mathsf{T}}$ of $\boldsymbol{\beta} = (\boldsymbol{\varphi}^{\mathsf{T}}, \boldsymbol{\theta}^{\mathsf{T}}, \sigma)^{\mathsf{T}}$.

If it can be safely assumed that the distribution of the ($\epsilon_t$, $t \in \mathbb{Z}$) generating the $Y_t$'s is of a given form, in particular independent identically distributed (*i.i.d.*) normal (Gaussian) random variables, then better inference can be obtained from the fitted model. For example, such an assumption is helpful to get accurate confidence or tolerance bounds for a predicted $Y_{T+h}$. Moreover, under this Gaussian assumption, $\hat{\boldsymbol{\beta}}$ is asymptotically efficient. It is thus important to have a tool to check the null hypothesis

$$H_0: \text{ the } \epsilon_t\text{'s are } i.i.d. \sim N(0, \sigma^2). \tag{2}$$

As pointed out by Pierce and Gray (1985) and Brockett et al. (1988), other reasons may motivate a test of (2). One such reason is to check the fit of the structural part of (1). Indeed, the process of fitting a model to data often reduces to finding the model whose residuals behave most like a sample of $i.i.d.$ Gaussian variables. In this context, rejection of (2) may indicate lack-of-fit of the entertained ARMA model. We will not elaborate further here on this possibility and assume, in the sequel, that model (1) is not underspecified. Note however that there exist specific tests for detecting lack-of-fit (for a recent review, see Koul and Stute (1999)).

For the problem of testing (2), the few tests available fall roughly into two groups. Tests of the first group use the fact that for the ARMA ($p$, $q$) models, normality of the $Y_t$'s induces normality of the $\epsilon_t$'s and vice versa. Thus a test of the hypothesis that a process ($Y_t$, $t \in \mathbb{Z}$) is Gaussian (Lomnicki (1961); Hinich (1982); Epps (1987)) can serve for problem (2). This presents the advantage of not requiring the values of $p$ and $q$. But Gasser (1975) and Granger (1976) have shown, and Lutkepohl and Schneider (1989) have confirmed by simulation, that this approach may lose much power. This is because the central limit theorem forces the $Y_t$'s to be close to normality even when (2) is false. Moreover, the adaptation of standard normality tests to dependent data is not an easy task. A small simulation study by Heuts and Rens (1986) has shown that, because of the serial correlation between the $Y_t$'s, the finite null behavior of standard normality tests based on the empirical distribution function (EDF) of the $Y_t$'s is different from what is obtained under $i.i.d.$ data. The same problem appears for tests based on the third or fourth moment of $Y_t$ (see Lomnicki (1961); Lutkepohl and Schneider (1989)) and for Pearson's chi-square test (Moore (1982)).

It thus appears better, when there are reasons to believe that a given ARMA($p$, $q$) model holds, to "inverse filter" the data and compute the residuals $\hat{\epsilon}_t$ of the fitted model. These can then be subjected to some test of normality. Tests of the second group are based on this idea and some examples are listed in Hipel and McLeod (1994). However, these and other authors use such tests in conjunction with critical values for $i.i.d.$ data. Since the residuals of an ARMA model are dependent, the null distribution of standard test statistics may be affected and critical values for $i.i.d.$ data may no longer be valid. It turns out that for AR models, there is theoretical evidence that this dependence affects only slightly the critical values, at least when $T$ is large. For an AR($p$)

model with unknown $E(Y_t)$, Pierce and Gray (1985) has shown that the asymptotic null distribution of any test statistic based on the EDF of the residuals coincides with that of the same statistic for $i.i.d.$ data with mean and variance unknown. Thus one can insert the residuals from an AR($p$) model into any of the standard EDF-based tests (Kolmogorov-Smirnov, Anderson-Darling) and if $T$ is large, use the critical values given, for example, in Chapter 4 of D'Agostino and Stephens (1986), to obtain an asymptotically valid test strategy. In the same vein, Lee and Na (2002) have recently adapted the Bickel-Rosenblatt test to this AR setting. Beiser (1985) has found that for the AR(1) model, tests based on the skewness or kurtosis coefficient of the residuals (D'Agostino and Stephens (1986), p. 408) in conjunction with the critical points derived for $i.i.d.$ data produce valid levels if $T$ is large and the AR-parameter is not too close to its boundary. This has been confirmed by Lutkepohl and Schneider (1989). See also Anděl (1997).

For the general ARMA model, much less is known. Ojeda et al. (1997) show that tests based on quadratic forms in differences between sample moments and expected values of certain non-linear functions of the sample have the same asymptotic distribution under the ARMA model as under $i.i.d.$ data. This suggests that a generalization of Pierce and Gray (1985) theorem to ARMA models could hold although, to our knowledge, no proof of this has been published. In accordance with this conjecture, the practice recommended in many textbooks (see for example, Brockwell and Davis (1991), p. 314; Hipel and McLeod (1994), p. 241) is to use standard normality tests in conjunction with critical values for $i.i.d.$ data.

In this paper, we develop some tests designed specifically for problem (2) in the ARMA($p$, $q$) context. Our approach is based on the smooth test paradigm introduced by Neyman (1937) and improved by the data driven technology introduced by Ledwina (1994) to select the best order for the test. This approach has been shown in the $i.i.d.$ case to offer many advantages, both theoretically and empirically, over other tests. In particular, the test statistic we recommend for problem (2) is easy to compute with an asymptotic $\chi^2$ distribution that can be corrected in finite samples to yield a close to nominal level. Moreover, as a byproduct of the procedure, diagnostic information is available that helps in understanding which aspects of the null hypothesis are not supported by the data.

Note that we concentrate here on the development of valid tests along this paradigm and do not dwell into their theoretical properties (i.e. local power and asymptotic efficiency). We also stress that the tests proposed here are valid solely for the case where $E(Y_t)$ is assumed known. The case where an unknown trend is present in (1) requires a special treatment and is the object of current research.

The paper is organized as follows. In Section 2, we develop the smooth goodness-of-fit test in the ARMA($p$, $q$) context of (1). In Section 3, we describe the data-driven technology that allows to "fine tune" the test by choosing a good value for its order. In Section 4, a Monte-Carlo study is conducted for some values of ($p$, $q$) to study the behavior of the proposed tests under the null hypothesis and compare their power to some competitors. It emerges that, under the null hypothesis, one of our data driven smooth tests holds its level over most of the parameter space and, under the alternatives studied, is in general more powerful than other methods. It can thus be recommended as a good tool for problem (2). An example concludes the paper.

## 2. SMOOTH TEST OF NORMALITY IN THE ARMA CONTEXT

Let $\Phi(.)$ be the cumulative distribution function of the $N(0,1)$ distribution with density $\phi(.)$ and let $U_t = 2\Phi(\epsilon_t/\sigma) - 1$ with density $g(.)$. Under $H_0$ of (2), the $U_t$'s are $i.i.d.$ U[-1, 1] so that (2) reduces to testing $g(u) = 1/2$ on [-1, 1]. The $\epsilon_t$'s are unobserved so the test must be based on residuals. Since the process $(Y_t, t \in \mathbb{Z})$ is invertible, we have

$$\epsilon_t = -\sum_{r=0}^{\infty} \delta_r Y_{t-r} \tag{1}$$

where the $\delta_r$'s are functions of $\boldsymbol{\theta}$ and $\boldsymbol{\varphi}$ (see (A.2), (A.3) of Appendix A). Let $\hat{\delta}_r$ be the Gaussian maximum likelihood estimator (*m.l.e.*) of $\delta_r$ under (2), obtained by plugging in the *m.l.e.* $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\varphi}}$ under $H_0$. We define the residuals of the fitted ARMA model by

$$\hat{\epsilon}_t = -\sum_{r=0}^{\infty} \hat{\delta}_r Y_{t-r}. \tag{2}$$

In practice, some scheme must be used to compute these residuals, for example by taking $Y_t = 0$ if $t < 1$. Note that other residuals can be defined for ARMA models (see Brockwell and Davis (1991), Section 9.4) but the definition above is convenient for the following derivation. Consider $\hat{U}_t = 2\Phi(\hat{\epsilon}_t/\hat{\sigma}) - 1$ , $t = 1, ..., T$. Let $\{L_k(.), k \geq 0\}$ be the normalized (over [-1, 1]) Legendre polynomials (Sansone (1959)) with $L_0(.) \equiv 1$ satisfying

$$\frac{1}{2}\int_{-1}^{1} L_k(x)L_j(x)dx = 1 \text{ if } k = j \text{ and } 0 \text{ otherwise.} \tag{3}$$

For some integer $K \geq 1$, consider the density defined on [-1, 1] by

$$g_K(u; \boldsymbol{\omega}) = c(\boldsymbol{\omega}) \exp\left\{\sum_{k=1}^{K} \omega_k L_k(u)\right\}, \tag{4}$$

where $c(\boldsymbol{\omega})$ is a normalizing constant such that $c(\mathbf{0}) = 1/2$. In the smooth test paradigm, (4) is the $K$-th order alternative with $g_K(.; 0)$ being the U[-1, 1] density. Thus, if $g(u)$ can be approximated by (4), (2) reduces to testing $H_0$: $\boldsymbol{\omega} = \mathbf{0}$. For this, we use the following route. Let $\boldsymbol{L}_t = (L_1(U_t), ..., L_K(U_t))^{\mathsf{T}}$, $\hat{\boldsymbol{L}}_t = (L_1(\hat{U}_t), ..., L_K(\hat{U}_t))^{\mathsf{T}}$ and

$$\overline{\hat{\boldsymbol{L}}} = T^{-1}\sum_{t=1}^{T} \hat{\boldsymbol{L}}_t. \tag{5}$$

Under $H_0$, $\boldsymbol{L}_t$ has mean $\mathbf{0}$ and covariance matrix $\boldsymbol{I}_K$, the $K$-th order identity matrix. Under (4), these moments will differ and (5) can be used to capture departures from the U[-1, 1] in the "direction" of $g_K(.; \boldsymbol{\omega})$. This suggests as a test statistic a quadratic form in $\overline{\hat{\boldsymbol{L}}}$ . To complete the test, we need the null asymptotic distribution of (5). This is given in the following theorem.

**Theorem 1.** *Consider the causal and invertible ARMA(p, q) process of (1) where we assume* $1 - \varphi_1 z - ... - \varphi_p z^p$ *and* $1 + \theta_1 z + ... + \theta_q z^q$ *have no common zeroes. Under* $H_0$*, we have*

$$\sqrt{T}\overline{\hat{\boldsymbol{L}}} \xrightarrow{L} N_K\left(\mathbf{0}, \boldsymbol{I}_K - \frac{1}{2}\boldsymbol{b}_K\boldsymbol{b}_K^{\mathsf{T}}\right) \tag{6}$$

*where* $\boldsymbol{b}_K = (b_1, ..., b_K)^{\mathsf{T}}$, *with* $b_k = \int_{\mathbb{R}} L_k(2\Phi(x) - 1)x^2\phi(x)dx$. *Hence, the smooth test statistic*

$$\mathcal{R}_K = T\overline{\hat{\boldsymbol{L}}}^{\mathsf{T}}\left(\boldsymbol{I}_K - \frac{1}{2}\boldsymbol{b}_K\boldsymbol{b}_K^{\mathsf{T}}\right)^{-1}\overline{\hat{\boldsymbol{L}}} \xrightarrow{L} \chi^2_K.$$

PROOF. We present an outline of the argument. More details are given in the appendices and in Ducharme and Lafaye de Micheaux (2002). Let

$$\mathcal{I}_{\boldsymbol{\beta}} = \mathrm{Var}\left[\frac{\partial}{\partial\boldsymbol{\beta}}\mathrm{Log}\left(\frac{1}{\sigma}\phi\left(\frac{\epsilon_t}{\sigma}\right)\right)\right]$$

be Fisher's information matrix for $\boldsymbol{\beta}$. From standard results (see Gouriéroux and Monfort (1995), p.325), we have,

$$\sqrt{T}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathcal{I}_{\boldsymbol{\beta}}^{-1}\frac{\partial}{\partial\boldsymbol{\beta}}\left[\mathrm{Log}\left(\frac{1}{\sigma}\phi\left(\frac{\epsilon_t}{\sigma}\right)\right)\right] + o_P(1).$$

Since $\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) = O_P(T^{-1/2})$, a Taylor expansion yields

$$\sqrt{T}\overline{\hat{\boldsymbol{L}}} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\boldsymbol{L}_t + \left[\frac{1}{T}\sum_{t=1}^{T}\frac{\partial}{\partial\boldsymbol{\beta}}\boldsymbol{L}_t\right]\sqrt{T}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) + o_P(1). \tag{7}$$

The first term on the right hand side of (7) converges to a $N_K(\mathbf{0}, \boldsymbol{I}_K)$. Moreover, it is shown in Appendix A that

$$\left[\frac{1}{T}\sum_{t=1}^{T}\frac{\partial}{\partial\boldsymbol{\beta}}\boldsymbol{L}_t\right] \xrightarrow{P} \left[\underset{K\times(p+q)}{\mathbf{0}}, -\frac{1}{\sigma}\boldsymbol{b}_K\right] = -\mathcal{J}_K. \tag{8}$$

Hence,

$$\begin{aligned}\sqrt{T}\overline{\hat{\boldsymbol{L}}} &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\boldsymbol{L}_t - \frac{1}{\sqrt{T}}\mathcal{J}_K\mathcal{I}_{\boldsymbol{\beta}}^{-1}\sum_{t=1}^{T}\frac{\partial}{\partial\boldsymbol{\beta}}\left[\mathrm{Log}\left(\frac{1}{\sigma}\phi\left(\frac{\epsilon_t}{\sigma}\right)\right)\right] + o_P(1)\\ &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\boldsymbol{B}\boldsymbol{V}_t + o_P(1),\end{aligned}$$

where $\boldsymbol{B} = (\boldsymbol{I}_K, -\mathcal{J}_K\mathcal{I}_{\boldsymbol{\beta}}^{-1})$ and

$$\boldsymbol{V}_t = \left(\boldsymbol{L}_t^{\mathsf{T}}, \frac{\partial}{\partial\boldsymbol{\beta}^{\mathsf{T}}}\left[\mathrm{Log}\left(\frac{1}{\sigma}\phi\left(\frac{\epsilon_t}{\sigma}\right)\right)\right]\right)^{\mathsf{T}}.$$

From Appendix B, it follows that, $E(\boldsymbol{V}_t) = \mathbf{0}$ and Var$(\boldsymbol{B}\boldsymbol{V}_t) = \boldsymbol{I}_K - \boldsymbol{b}_K\boldsymbol{b}_K^{\mathsf{T}}/2$ . The central limit theorem yields (6). $\square$

It is possible to write $\mathcal{R}_K$ in a form that makes it easy to use. A Cholesky decomposition of $\left(\boldsymbol{I}_K - \boldsymbol{b}_K\boldsymbol{b}_K^\mathsf{T}/2\right)$ yields $\left(\boldsymbol{I}_K - \boldsymbol{b}_K\boldsymbol{b}_K^\mathsf{T}/2\right)^{-1} = \boldsymbol{P}\boldsymbol{P}^\mathsf{T}$ with $\boldsymbol{P} = (p_{ij})$, an upper triangular matrix. Some algebra gives $p_{ij} = 0$ if $i > j$, while

$$p_{ii} = \sqrt{\frac{2 - \sum_{k=1}^{i-1} b_k^2}{2 - \sum_{k=1}^{i} b_k^2}} \text{ and } p_{ij} = \frac{b_i b_j}{\sqrt{\left(2 - \sum_{k=1}^{j-1} b_k^2\right)\left(2 - \sum_{k=1}^{j} b_k^2\right)}} \text{ if } j > i.$$

Thus

$$\mathcal{R}_K = \sum_{k=1}^{K} \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} L_k^*(\hat{U}_t) \right)^2,$$

where

$$L_k^*(\hat{U}_t) = \sum_{l=1}^{k} p_{lk} L_l(\hat{U}_t). \tag{9}$$

Numerical integration gives $(b_2, b_4, ..., b_{10}) = (1.23281, 0.521125, 0.304514, 0.205589, 0.150771)$ with $b_k = 0$ if $k$ is odd. This yields the first ten "modified" Legendre polynomials

$L_1^*(u) = 1.73u,$
$L_2^*(u) = 6.85u^2 - 2.28,$
$L_3^*(u) = 6.61u^3 - 3.97u,$
$L_4^*(u) = 19.91u^4 - 10.26u^2 - 0.56,$
$L_5^*(u) = 26.12u^5 - 29.02u^3 + 6.22u,$
$L_6^*(u) = 69.84u^6 - 81.84u^4 + 28.36u^2 - 3.06,$
$L_7^*(u) = 103.84u^7 - 167.75u^5 + 76.25u^3 - 8.47u,$
$L_8^*(u) = 260.07u^8 - 450.18u^6 + 247.18u^4 - 38.73u^2 - 1.11,$
$L_9^*(u) = 413.92u^9 - 876.55u^7 + 613.58u^5 - 157.33u^3 + 10.73u,$
$L_{10}^*(u) = 994.51u^{10} - 2250.43u^8 + 1782.83u^6 - 569.92u^4 + 67.54u^2 - 3.58.$

**Remark 2.1.** *Theorem 1 shows that we can slightly extend the result of Pierce and Gray (1985) and state that neither the estimation of $\boldsymbol{\varphi}$ and $\boldsymbol{\theta}$ nor the dependence of the $Y_t$'s has any asymptotic impact on a smooth test of (2) in the ARMA context. In pre-asymptotic situations, these elements and the complexity of the model will affect the null distribution of $\mathcal{R}_K$. This will be further explored in simulations of Section 4.*

**Remark 2.2.** *Each term $T^{-1}\left(\sum_{t=1}^{T} L_k^*(\hat{U}_t)\right)^2$ is a component of the test statistic and has an asymptotic $\chi_1^2$ distribution under $H_0$. When the null hypothesis is rejected, some of these components will be large. The simple structure of the first few polynomials in (9) helps in understanding what aspects of the normal are not supported by the data. For example, the first component detects departure from symmetry under $H_0$ in the "direction" of asymmetry. This diagnostic analysis must be undertaken with some care however; see Henze (1997) for details.*

**Remark 2.3.** *The above methodology can in principle be applied to other distributions than the normal. For location-scale densities, one needs to replace the normal distribution in the definition of $U_t$ and follow the derivation using the new null density.*

*The structure of $\mathcal{R}_K$ will be similar to what is obtained above but the modified Legendre polynomials will change. For distributions with a shape parameter, the statistic is more complex since the coefficients of these polynomials will in general depend on this unknown shape parameter that must be estimated.*

## 3. CHOOSING THE ORDER $K$ OF THE ALTERNATIVE

Before applying the test strategy of Section 2, one must choose the value of $K$. Ideally, this choice should be made so that members of the embedding family $g_K(.;\boldsymbol{\omega})$ of (4) provide a good approximation to any plausible density $g(.)$ of $U_t$ under the alternative. If $K$ is too small, this approximation may be crude and the test loses power. If $K$ is too large, power dilution can occur since $g_K(.;\boldsymbol{\omega})$ encompasses unnecessary "directions".

In practice, the user has only, at best, a qualitative idea of the plausible alternatives and no specific value of $K$ emerges naturally. In the $i.i.d.$ case, some authors (Rayner and Best (1989)) argue that, as a rule of thumb, one can use a trade-off value of $K$ between 2 and 4.

Recently, Ledwina (1994) and Kallenberg and Ledwina (1997a,b) have proposed and explored for $i.i.d.$ data a method to choose adaptively a value for $K$. At the first step, Schwarz (1978)'s criterion is used to choose the value $\hat{K}$ that seems best in view of the data at hand. The smooth test strategy is then applied using the statistic $\mathcal{R}_{\hat{K}}$. Extensive simulations have shown that, even for small sample sizes, this so-called "data driven smooth test" can yield power close to what could be obtained if one knew the true form of the alternative and had chosen the best value of $K$ accordingly.

So far, this approach has been investigated for $i.i.d.$ data only but it can be extended to the ARMA context. Choose two integers $1 \leq d \leq D$ and consider the set of statistics $(\mathcal{R}_d, ..., \mathcal{R}_D)$. We seek a rule that will select a good $\mathcal{R}_K$ in this set. Write

$$\hat{K} = \min \left[ \underset{d \leq s \leq D}{\text{Argmax}} \{ \mathcal{R}_s - s\text{Log}(T) \} \right] \tag{1}$$

and denote $\mathcal{R}_{\hat{K}}(d)$, the test statistic $\mathcal{R}_{\hat{K}}$ selected by (1) in $(\mathcal{R}_d, ..., \mathcal{R}_D)$.

**Theorem 2.** *Under $H_0$, $\hat{K} \longrightarrow d$ in probability and thus, $\mathcal{R}_{\hat{K}}(d)$ is asymptotically $\chi^2_d$.*

PROOF. Set $e_k = (k-d)\text{Log}T$. For $k \geq d$, $P(\hat{K} = k) \leq P(\mathcal{R}_k > e_k)$. Now, since each $\mathcal{R}_k$ is asymptotically $\chi^2_k$ under $H_0$, as $T$ increases,

$$P(\mathcal{R}_k > e_k) \longrightarrow 0,$$

when k > d. It follows that $P(\hat{K} = d) = 1 - P(\hat{K} \geq d+1) \longrightarrow 1$. □

For finite sample sizes, the asymptotic null distribution of Theorem 2 may not provide a good approximation to that of $\mathcal{R}_{\hat{K}}(d)$ since there is a positive probability that $\hat{K} \geq d+1$. A simple correction has been developed by Janic-Wroblewska and Ledwina (2000) when $d = 1$ ($i.i.d.$ data). Because of the asymptotic independence between the components of $\mathcal{R}_k$, this correction can easily be extended to $d > 1$ and to the present

ARMA context. A direct application of the argument in their Section 4 leads to the following approximation, which can be solved for $x$ by numerical integration

$$P(\mathcal{R}_{\hat{K}}(d) \leq x) \approx P(\chi_d^2 \leq x)P(\chi_1^2 \leq \text{Log}(T)) + \int_{\text{Log}(T)}^{x} P(\chi_d^2 < x - z)\frac{1}{\sqrt{2\pi z}}e^{-z/2}dz. \tag{2}$$

Some quantiles corrected through (2) are listed in Table 3.1.

TABLE 3.1. Some quantiles obtained from approximation (2)

| | T | a = 0.10 | a = 0.05 | a = 0.01 |
|---|---|---|---|---|
| d = 1 | 50 | 3.692 | 5.410 | 8.805 |
| | 100 | 3.275 | 5.201 | 8.703 |
| | 200 | 3.057 | 4.751 | 8.590 |
| d = 2 | 50 | 5.466 | 7.137 | 10.807 |
| | 100 | 5.262 | 6.972 | 10.684 |
| | 200 | 5.043 | 6.796 | 10.558 |

One may have the feeling that this data driven approach replaces the problem of selecting $K$ with that of selecting $d$ and $D$. To answer this, Kallenberg and Ledwina (1997a,b) have studied a version of the above procedure where $D$ is allowed to increase with $T$. In the $i.i.d.$ case, they obtain rates connecting these quantities. These rates are theoretically interesting but do not help in practice in selecting a value for $D$. To get more insight, they have conducted extensive simulations. It turns out that the power levels off rapidly as $D$ increases and there is little to be gained by choosing $D$ much greater than 10. As for the choice of $d$, again Kallenberg and Ledwina (1997a) briefly discuss this problem where it emerges that in their context $d = 1$ or 2 appears reasonable. In the simulation study of the next section we use both these values of $d$ and take $D = 10$.

In closing this section, note that, by plotting $g_{\hat{K}}(.;\hat{\boldsymbol{\omega}})$ where $\hat{\boldsymbol{\omega}}$ is an estimate of $\boldsymbol{\omega}$, one can get an idea of the true shape of the density when the null hypothesis has been rejected. This can be helpful in finding a more appropriate distribution for the innovations.

## 4. SIMULATION RESULTS

To get an idea of the behavior of our test statistics as compared to some competitors, a simulation study was conducted. Samples $\{Y_t, t = 1, ..., T\}$ from various ARMA($p$, $q$) models were generated with the innovations arising, in the first part of the simulation, from the normal distribution and, in the second, from various alternatives. For each sample, we estimated the parameters of the model and computed test statistics. From there, we obtained approximations to their level and power. All programs are written in Fortran 77. The subroutines listed below are from the Numerical Algorithms Group (NAG) MARK 16 Fortran library.

### 4.1. Levels

The first part of the simulation study was designed to see if the critical values obtained from the asymptotic $\chi^2$ or from (2) can be relied upon in finite samples. We took $T = 50$, 100 and 200 and restricted attention to the models MA(2), AR(2), ARMA(1,2), ARMA(2,1) and ARMA(2, 2). To generate ARMA($p$,$q$) samples with Gaussian innovations, we used subroutine G05EGF and G05EWF. These samples were submitted to subroutine G13DCF that returns estimates of the parameters of the model as well as residuals. The definition of these residuals, given at equation (9.4.1) in Brockett et al. (1988), differs from (2) but their numerical values are almost identical. These residuals were then submitted to the various tests. The actual level of each test was computed for nominal level $\alpha = 0.10$ and $0.05$.

Regarding the parameter $\boldsymbol{\beta}$, note that our test statistics are in theory invariant to the choice of $\sigma$ and we took $\sigma = 1$. Numerically, this invariance holds approximately because of the stopping rule in G13DCF. But the finite distribution of our test statistics depends on the values of $\boldsymbol{\theta}$ and $\boldsymbol{\varphi}$. To explore this, we have proceeded as follows. First, causality requires that, if $p = 1$, $\varphi_1 \in ]-1, 1[$ while if $p = 2$, $\boldsymbol{\varphi}$ must be in the region $\triangle_{\boldsymbol{\varphi}} = \{(\varphi_1, \varphi_2)|\varphi_1 + \varphi_2 < 1, \varphi_2 - \varphi_1 < 1, |\varphi_2| < 1\}$ (Brockett et al. (1988), p. 110, ex.3.2). Similarly, invertibility implies that if $q = 1$, $\theta_1 \in ]-1, 1[$ while if $q = 2$, $\boldsymbol{\theta}$ must be in $\bigtriangledown_{\boldsymbol{\theta}} = \{(-\theta_1, -\theta_2)|\theta_1 + \theta_2 < 1, \theta_2 - \theta_1 < 1$ and $|\theta_2| < 1\}$. In addition, the polynomials $1 - \varphi_1 z$ when $p = 1$ and $1 - \varphi_1 z - \varphi_2 z^2$ when $p = 2$ must have no common zeroes with $1 + \theta_1 z$ when q = 1 and $1 + \theta_1 z + \theta_2 z^2$ when q = 2.

For the AR(2) model, we have taken the values of $\boldsymbol{\varphi}$ in the grid of 64 points $\{(-2.0 + 0.25j, -0.9 + 0.25k) \in \triangle_{\boldsymbol{\varphi}}|j, k \geq 0\}$. A similar grid was used for the MA(2). This makes it possible to see whether the tests maintain the proper critical level over a large section of the parameter space. For the ARMA(1, 2), the grid over $\bigtriangledown_{\boldsymbol{\theta}}$ was reduced to $\{(-2.0 + 0.40j, -0.9 + 0.40k) \in \bigtriangledown_{\boldsymbol{\theta}}|j, k \geq 0\}$ while $\varphi_1 = -0.9 + 0.2j, j = 0, ..., 9$. This gives a set of 250 points on the parameter space of $(\varphi_1, \boldsymbol{\theta})$. For the ARMA(2, 1) model, the same was done with $\boldsymbol{\varphi}$ and $\theta_1$ instead. Finally, for the ARMA (2, 2) model, points $(\boldsymbol{\varphi}, \boldsymbol{\theta})$ satisfying the "no common zeroes" condition were taken in $\{(-1.95 + 0.45j, -0.85 + 0.45k) \in \triangle_{\boldsymbol{\varphi}}|j, k \geq 0\} \cup \{-(-1.95 + 0.45j, -0.95 + 0.45k) \in \bigtriangledown_{\boldsymbol{\theta}}|j, k \geq 0\}$. This yields 294 $(\boldsymbol{\varphi}, \boldsymbol{\theta})$ parameter points. For each of these parameter points, 10000 samples of size $T$ were generated as described above.

To summarize the results, the following approach was adopted. A 95% confidence interval for the true level when $\alpha = 0.10$ is (0.094, 0.106). Similarly, for $\alpha = 0.05$, 95% of the $p$-values are expected in the interval (0.046, 0.054). Thus the range of possible $p$-values was divided in 5 sub-intervals. For $\alpha = 10\%$, these are $I_1 = (0, 0.085)$, $I_2 = [0.085, 0.094)$, $I_3 = [0.094, 0.106)$, $I_4 = [0.106, 0.115)$ and $I_5 = [0.115, 1]$. For $\alpha = 0.05$, $I_1 = (0, 0.035)$, $I_2 = [0.035, 0.046)$, $I_3 = [0.046, 0.054)$, $I_4 = [0.054, 0.065)$ and $I_5 = [0.065, 1]$. For each model, the percentage of $p$-values in each interval was recorded. Table 4.1 reports the results for statistics $\mathcal{R}_3$ and $\mathcal{R}_{\hat{K}}(2)$ which, as discussed in Section 3, are representative of the two schools of thought for the choice of $K$. The results for the AR(2) and ARMA(2,1) models being similar to those of the MA(2) and ARMA(1,2) respectively, are omitted for brevity (see Ducharme and Lafaye de Micheaux (2002) for more complete results).

The actual levels for $\mathcal{R}_3$ are concentrated on $I_1$, $I_2$ and $I_3$. The mode of the distribution is generally located on $I_2$ for $T = 50$ and is shifted to $I_3$ as $T$ increases. This lead, at worst, to slightly conservative tests. To appreciate this, the last column of Table 4.1 gives the smallest $p$-value recorded over the parameter points. For $\mathcal{R}_{\hat{K}}(2)$, the distribution is concentrated on $I_2$, $I_3$ and $I_4$ with, in all cases, a mode centered on $I_3$. For this statistic, the minimal $p$-values are also closer to the nominal level (no maximal $p$-value was very far from the upper bound of $I_4$). Thus correction (2) works nicely, at least for the cases considered here.

We also investigated what areas of the parameter space give $p$-values in $I_1$. Intuitively, one expects these points to be near the boundary. However, the pattern that emerges, which is very similar for both $\mathcal{R}_3$, and $\mathcal{R}_{\hat{K}}(2)$, is more precise. For AR(2) models, these points correspond mainly to positive $(\varphi_1, \varphi_2)$ close to the right boundary of $\triangle_{\boldsymbol{\varphi}}$ and, to a lesser degree, to those with positive $\varphi_1$ and negative $\varphi_2$ but again close to that boundary. For MA(2) models, the situation is reversed, which is not surprising since $\bigtriangledown_{\boldsymbol{\theta}} = -\triangle_{\boldsymbol{\varphi}}$. For ARMA(2, 1), the points giving small $p$-values correspond to positive $(\varphi_1, \varphi_2)$ combined with values of $\theta_1$ close to -1. Again, for ARMA(1, 2) the situation is reversed and small $p$-values are associated with negative values of $(\theta_1, \theta_2)$ with a value of $\varphi_1$ close to 1. Finally, for the ARMA(2, 2), the points that yield $p$-values in $I_1$ are mainly those with positive $(\varphi_1, \varphi_2)$ and negative $(\theta_1, \theta_2)$.

We have also investigated the behavior under $H_0$ of some other tests that have been recommended in the time series literature for (2). We first considered the Anderson-Darling ($\mathcal{AD}$) test (Pierce and Gray (1985)) for case 2 (known mean) used in conjunction with the quantiles given in D'Agostino and Stephens (1986) p. 122. Our simulations show that, for large $T$ this yields valid critical levels. We also studied a variant of the Shapiro-Wilk test known as the Weisberg and Bingham (1975) ($\mathcal{WB}$) test. To adapt this test to our context where the mean is known, the denominator of equation (9.68) of D'Agostino and Stephens (1986) was replaced by $T\hat{\sigma}^2$ , where $\hat{\sigma}^2$ is the estimate of $\sigma^2$ returned by subroutine G13DCF. Up to the numerical accuracy of procedure G13DCF, this corresponds to the sum of squares of the residuals. Our simulations show that the quantiles for this test can be approximated by Monte Carlo using $i.i.d.$ data, although we found no theoretical result supporting this. Thus, we simulated 100000 samples from an ARMA(0,0) model and computed the empirical quantiles. For $T$ = 50, 100 and 200, we got, for $\alpha = 10\%$, 0.920, 0.958 and 0.978. For 5%, we found 0.899, 0.947 and 0.973. A third approach, the Jarque and Bera (1987) eq. (5) ($\mathcal{JB}$) test was also investigated. Although developed in the linear regression context, this test has been recommended in the time series literature (see Cromwell et al. (1994); Frances (1998)). A summary of the results for these tests in the ARMA(1, 2) model is given in Table 4.2. Also appearing in this table are the levels of the test based on $\mathcal{R}_{\hat{K}}(1)$ using quantiles derived from (refequation3.2).

Overall, the best tests, according to the criterion of maintaining the proper level throughout the parameter space, are $\mathcal{R}_{\hat{K}}(2)$ followed by $\mathcal{R}_{\hat{K}}(1)$ and then $\mathcal{R}_3$, $\mathcal{AD}$ and

$\mathcal{WB}$. In general, the AD test yields distributions of $p$-values in between those of $\mathcal{R}_3$ and $R_{\hat{K}}(1)$. More troublesome is the fact that this test, as well as the $\mathcal{WB}$ test, may vastly underestimate the intended level, as can be seen by the minimal $p$-values (last column of Table 4.2) encountered on the grids. Also, there appears to be a problem with the $\mathcal{JB}$ test as the quantiles, obtained from the $\chi_2^2$ approximation, lead to gross error. Further simulations indicate that the convergence to the $\chi_2^2$ is very slow. The $\mathcal{JB}$ statistic is a version of the Bowman and Shenton test statistic that, for $i.i.d.$ data, has a notoriously slow convergence. The simulation results in Lutkepohl and Schneider (1989) tend to show that this is also the case for AR(1) and AR(2) models. In view of this problem, we choose to drop from further investigations the $\mathcal{JB}$ test.

### 4.2. Power

The second part of the simulation was designed to study the power of our tests and allow comparison with the competitors mentioned above. We restricted attention to $i.i.d.$ innovations. We generated samples $\{Y_t, t = 1, ..., T\}$ according to model (1) from various alternatives to the normal distribution. These alternatives were taken as the centered version of the densities listed in Table V of Kallenberg and Ledwina (1997b). They comprise a large range of departure from the normal distribution both in skewness, kurtosis and shape.

To generate ARMA($p$, $q$) samples $\{Y_t, t = 1, ..., T\}$ according to model (1) with non-Gaussian innovations, we used the random shock method (algorithms IA 1 with $m = 50$ and SA 1 with $M = 200$) of Burn (1987). To allow a proper comparison of the various tests, we used for each model a set of parameters for which the $p$-values computed in the first part of the simulation were in $I_3$ for all tests. More precisely we took: ARMA(2, 1): $(\boldsymbol{\varphi}, \theta_1) = (-0.8, -0.1, 0.7)$, ARMA(1, 2): $(\varphi_1, \boldsymbol{\theta}) = (-0.7, 0.4, 0.5)$ and ARMA(2, 2): $(\boldsymbol{\varphi}, \boldsymbol{\theta}) = (-1.05, -0.4, 0.15, 0.85)$. Also we took $T = 50$ (more complete simulations appear in Ducharme and Lafaye de Micheaux (2002)). For each combination of model and alternative distribution, we generated 10000 samples and performed the various tests. From there, empirical powers were computed.

Table 4.3 presents these empirical powers for the tests $\mathcal{R}_3$, $\mathcal{R}_{\hat{K}}(2)$ and $\mathcal{WB}$ when $\alpha = 10\%$. Similar results were obtained for $\alpha = 5\%$. The tests $\mathcal{R}_3$ and $\mathcal{R}_{\hat{K}}(2)$ behave similarly with, overall, $\mathcal{R}_{\hat{K}}(2)$ being slightly better. Both these tests generally dominate the others. The $\mathcal{AD}$ approach, not shown here, often yields a power that is much lower than these two tests whereas $\mathcal{WB}$ generally lies somewhere in between. For $i.i.d.$ data, the $\mathcal{WB}$ test, as a variant of the Shapiro-Wilk test, is considered among the best omnibus tests of normality. In ARMA situations, this does not seem to hold at the same degree.

We have also computed the power of the test based on $\mathcal{R}_{\hat{K}}(1)$. The tabulated results are not presented here for brevity. We found that, for $T = 50$ and symmetric alternatives, the test based on $\mathcal{R}_{\hat{K}}(1)$ yields slightly better power that $\mathcal{R}_{\hat{K}}(2)$. For asymmetric alternatives, the situation is reversed. But for $T = 100$, $\mathcal{R}_{\hat{K}}(2)$ is more powerful almost everywhere. This behavior of $\mathcal{R}_{\hat{K}}(1)$ is explained by the fact that for asymmetric alternatives, $\mathcal{R}_1$ yields little, sometimes trivial, power. Moreover, power as a function of $K$ usually levels off at $K = 3$, and not infrequently at $K = 2$. This empirical observation is behind the rule of thumb stated in Section 3. Thus to have good power, the selection rule with $d = 1$ must give $\hat{K} \geq 3$, which may be difficult. Starting at $d = 2$ gives a better chance that $\hat{K} \geq 3$ when necessary.

In view of the results of these simulations, we recommend the use of $\mathcal{R}_{\hat{K}}(2)$ for testing (2) when $E(Y_t)$ is known. The levels are stable over most of the parameter points and close to nominal for moderate samples. Moreover, the power is generally better than that of other tests that have been recommended in the time series literature. Finally, the test is very easy to apply.

## 5. AN EXAMPLE

In the course of a study to forecast the amount of daily gas required, Shea (1987) has studied a bivariate time series of $T = 366$ points. The first component of this time series pertains to differences in daily temperature between successive days ($\bigtriangledown\tau_t$) and he found, after an iteration process of fitting and diagnostic checking, that the following MA(4) model could be entertained:

$$\bigtriangledown\tau_t = \epsilon_t + 0.07\epsilon_{t-1} - 0.30\epsilon_{t-2} - 0.15\epsilon_{t-3} - 0.20\epsilon_{t-4}.$$

The residual variance is 2.475. All these parameters are obtained by maximizing the Gaussian likelihood so that problem (2) is of some importance. Shea does not discuss the normality of the innovations in assessing the fit of this model but rather goes on to find a good model for the bivariate series based on an analysis of the residuals' cross correlation matrix.

An application of our tests yields $\mathcal{R}_3 = 22.85$, with a $p$-value of 0.00004 while $\mathcal{R}_{\hat{K}}(2) = 22.77$ ($\hat{K} = 2$) yielding a $p$-value of 0.00003 according to (2). Thus, both tests strongly reject the null hypothesis (2). A complementary analysis helps understanding what aspect of the Gaussian is not supported by the data. We found $\mathcal{R}_1 = 0.15$ ($p = 0.69$) with a skewness coefficient of 0.13. Thus there is no reason to suspect an asymmetrical distribution for the innovations. On the other hand, we can notice that 9.3% of the absolute standardized residuals are greater than 2.5 and the kurtosis is 4.33. Thus, if the model entertained above is correct, the conclusion that emerges from the present analysis is that the $\bigtriangledown\tau_t$ series could have been generated from innovations with a symmetric distribution having fatter tails than the Gaussian.

**ACKNOWLEDGMENTS**
The authors would like to thank Dr. B.L. Shea for some insight on subroutine G13DCF of the NAG library and for providing them with the data set used in Section 5.

## APPENDIX A

We show that(8) holds under $H_0$. Assume $p$ and $q > 0$. It suffices to show that

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\partial}{\partial\sigma}L_k(U_t) \xrightarrow{P} E\left[\frac{\partial}{\partial\sigma}L_k(U_t)\right] = -\frac{1}{\sigma}b_k, \tag{A.1.a}$$

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\partial}{\partial\varphi_1}L_k(U_t) \xrightarrow{P} E\left[\frac{\partial}{\partial\varphi_1}L_k(U_t)\right] = 0, \tag{A.1.b}$$

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\partial}{\partial\theta_1}L_k(U_t) \xrightarrow{P} E\left[\frac{\partial}{\partial\theta_1}L_k(U_t)\right] = 0. \tag{A.1.c}$$

First,

$$\frac{\partial}{\partial\sigma}L_k(U_t) = -\frac{2\epsilon_t}{\sigma^2}\phi\left(\frac{\epsilon_t}{\sigma}\right)L_k'(x)|_{x=2\Phi\left(\frac{\epsilon_t}{\sigma}\right)-1} = -\frac{\epsilon_t}{\sigma^2}w\left(\frac{\epsilon_t}{\sigma}\right) \text{ say.}$$

The law of large numbers yields (A.1.a). For (A.1.b), define for $r \geq 0$,

$$B_{r-1} = \frac{\partial}{\partial\varphi_1}\delta_r(\boldsymbol{\theta},\boldsymbol{\varphi}),$$

where, setting $\varphi_0 = -1$, $\gamma_0 = \theta_0 = 1$, we have

$$\delta_r(\boldsymbol{\theta},\boldsymbol{\varphi}) = \delta_r = \sum_{i=0}^{\min(r,p)}\varphi_i\gamma_{r-i} \quad r \geq 0, \tag{A.2}$$

$$\gamma_r = -\sum_{i=1}^{\min(r,q)}\gamma_{r-i}\theta_i \quad r \geq 1. \tag{A.3}$$

Obviously $B_{r-1} = \gamma_{r-1}$ when $r \geq 1$. For $r \geq q$, from Brockwell and Davis (1991), p.107,

$$\gamma_r = \sum_{i=1}^{j}\sum_{n=0}^{r_i-1}c_{in}r^n\alpha_i^{-r}$$

for some constants $c_{in}$ and where the $\alpha_i$'s are the $j$ distinct roots of $1+\theta_1 z+...+\theta_q z^q$ and $r_i$ is the multiplicity of $\alpha_i$, $i = 1,...,j$. Thus, when $r \geq q+1$,

$$B_{r-1} = \sum_{i=1}^{j}\sum_{n=0}^{r_i-1}c_{in}(r-1)^n\alpha_i^{-r+1}. \tag{A.4}$$

If $(X_t, t\in\mathbb{Z})$ is a weak stationary process such that Cov$(X_t, X_{t+h}) \to 0$ as $h \to \infty$, then $\overline{X}_T \xrightarrow{P} E(X_t)$. We apply this result with $X_t = \partial L_k(U_t)/\partial\varphi_1$. From (1) and (1), we have

$$X_t = \frac{1}{\sigma}w\left(\frac{\epsilon_t}{\sigma}\right)\frac{\partial}{\partial\varphi_1}\epsilon_t = -\frac{1}{\sigma}w\left(\frac{\epsilon_t}{\sigma}\right)\left(Y_{t-1} - \sum_{r=0}^{\infty}\left(\frac{\partial}{\partial\varphi_1}\delta_r\right)\boldsymbol{\theta}^{\mathsf{T}}\mathbf{Y}_{t-1-r}^{(q)}\right). \tag{A.5}$$

Thus $E(X_t) = 0$. Moreover, Var$(X_t) < \infty$ as shown in Appendix C and it is seen by Lemma C.1 that Cov$(X_t, X_{t+h})$ depends only on $h$. Thus $(X_t, t\in\mathbb{Z})$ is stationary. We show that $\mathrm{Cov}(X_t, X_{t+h}) \to 0$ as $h \to \infty$. From (A.5), for $h$ large, $|Cov(X_t, X_{t+h})| = |d_1|E|w(\epsilon_{t+h}/\sigma)|/\sigma$, where

$$d_1 = E\left[\frac{1}{\sigma}w\left(\frac{\epsilon_t}{\sigma}\right)\left\{Y_{t-1} - \sum_{r=0}^{\infty} B_{r-1}\boldsymbol{\theta}^{\mathsf{T}}\boldsymbol{Y}^{(q)}_{t-1-r}\right\}\left\{Y_{t+h-1} - \sum_{r=0}^{\infty} B_{r-1}\boldsymbol{\theta}^{\mathsf{T}}\boldsymbol{Y}^{(q)}_{t+h-1-r}\right\}\right]. \tag{A.6}$$

But, $|d_1| \le d_2 + \sum_{j=1}^{q} |\theta_j|(d_{3j} + d_{4j}) + \sum_{i=1}^{q}\sum_{j=1}^{q} |\theta_i\theta_j|d_{5ij}$ where

$$d_2 = \left|E\left[\frac{1}{\sigma}w\left(\frac{\epsilon_t}{\sigma}\right)Y_{t-1}Y_{t+h-1}\right]\right|,\; d_{3j} = \sum_{r=0}^{\infty}\left|B_{r-1}E\left[\frac{1}{\sigma}w\left(\frac{\epsilon_t}{\sigma}\right)Y_{t-r-j}Y_{t+h-1}\right]\right|,$$

$$d_{4j} = \sum_{r=0}^{\infty}\left|B_{r-1}E\left[\frac{1}{\sigma}w\left(\frac{\epsilon_t}{\sigma}\right)Y_{t-1}Y_{t+h-r-j}\right]\right|$$

and

$$d_{5ij} = \sum_{r=0}^{\infty}\sum_{r'=1}^{\infty}\left|B_{r-1}B_{r'-1}E\left[\frac{1}{\sigma}w\left(\frac{\epsilon_t}{\sigma}\right)Y_{t-r-i}Y_{t+h-r'-j}\right]\right|.$$

It can be shown that $d_2$, $d_{3j}$, $d_{4j}$ and $d_{5ij} \to 0$ when $h \to \infty$. Proof for $d_{4j}$, which is typical, is sketched in Appendix D. This yields (A.1.b).

As for (A.1.c), let $A_r = \frac{\partial}{\partial\theta_1}\delta_r(\boldsymbol{\theta}, \boldsymbol{\varphi})$. From (A.3), we obtain, for $r \ge q$, the system

$$\begin{cases} \gamma'_r + \theta_1\gamma'_{r-1} + ... + \theta_q\gamma'_{r-q} = -\gamma_{r-1} \\ \gamma_r + \theta_1\gamma_{r-1} + ... + \theta_q\gamma_{r-q} = 0 \end{cases}$$

from which we find

$$0 = \sum_{j=0}^{q}\theta_j\left(-\sum_{i=0}^{q}\theta_i\gamma'_{r-j-i+1}\right) = \sum_{h=0}^{2q} a_h\gamma'_{r-h+1} \text{ where } a_h = \sum_{\substack{i+j=h\\0\le i,j\le q}}\theta_i\theta_j,\ \forall r \ge 2q-1.$$

Again from Brockwell and Davis (1991), p.107, we have, for some constants $d_{in}$

$$\gamma'_r = \sum_{i=1}^{j}\sum_{n=0}^{s_i-1} d_{in}r^n\beta_i^{-r}$$

where the $\beta_i$'s are the $j$ distinct roots (with multiplicity $s_i$) of $1+a_1z+a_2z^2+...+a_{2q}z^{2q}$. Now

$$\left(\sum_{i=0}^{q}\theta_i z\right)^2 = \sum_{h=0}^{2q}\left(\sum_{\substack{i+j=h\\0\le i,j\le q}}\theta_i\theta_j\right)z^h = \sum_{h=0}^{2q} a_h z^h$$

where $a_0 = \theta_0^2 = 1$. This shows that the roots of $1 + a_1z + a_2z^2 + ... + a_{2q}z^{2q}$ are exactly the same than that of $1 + \theta_1z + \theta_2z^2 + ... + \theta_qz^q$, apart from the multiplicity. Thus, we obtain

$$A_r = \sum_{l=0}^{p}\sum_{i=1}^{j}\sum_{n=0}^{s_i-1} d_{in}(r-l)^n\alpha_i^{-(r-l)}\varphi_l, \text{ for all } r \ge \max(2q, p). \tag{A.7}$$

By the same argument, using $A_r$ of (A.7) instead of $B_{r-1}$ of (A.4), we get (A.1.c).

## APPENDIX B

We show that $E(\boldsymbol{V}_t) = 0$ and Var$(\boldsymbol{B}\boldsymbol{V}_t) = \boldsymbol{I}_K - \boldsymbol{b}_K\boldsymbol{b}_K^{\mathsf{T}}/2$. In view of (1) and (1),

$$\frac{\partial}{\partial\boldsymbol{\varphi}}\text{Log}\left(\frac{1}{\sigma}\phi\left(\frac{\epsilon_t}{\sigma}\right)\right) = \frac{\epsilon_t}{\sigma^2}\left[\boldsymbol{Y}_{t-1}^{(p)} - \sum_{r=0}^{\infty}\left(\frac{\partial}{\partial\boldsymbol{\varphi}}\delta_r\right)\boldsymbol{\theta}^{\mathsf{T}}\boldsymbol{Y}_{t-1-r}^{(q)}\right],$$

$$\frac{\partial}{\partial\boldsymbol{\theta}}\text{Log}\left(\frac{1}{\sigma}\phi\left(\frac{\epsilon_t}{\sigma}\right)\right) = \frac{\epsilon_t}{\sigma^2}\left[\boldsymbol{\epsilon}_{t-1}^{(q)} - \sum_{r=0}^{\infty}\left(\frac{\partial}{\partial\boldsymbol{\theta}}\delta_r\right)\boldsymbol{\theta}^{\mathsf{T}}\boldsymbol{Y}_{t-1-r}^{(q)}\right]$$

and

$$\frac{\partial}{\partial\sigma}\text{Log}\left(\frac{1}{\sigma}\phi\left(\frac{\epsilon_t}{\sigma}\right)\right) = \frac{1}{\sigma}\left(\left(\frac{\epsilon_t}{\sigma}\right)^2 - 1\right).$$

It follows that $E(\boldsymbol{V}_t) = 0$ under $H_0$. Moreover, under $H_0$, Var$(\boldsymbol{L}_t) = \boldsymbol{I}_K$. Thus,

$$\text{Cov}\left(\boldsymbol{L}_t, \frac{\partial}{\partial\boldsymbol{\beta}}\text{Log}\left(\frac{1}{\sigma}\phi\left(\frac{\epsilon_t}{\sigma}\right)\right)\right)^{\mathsf{T}} = \left[\begin{array}{c}\mathbf{0}\\ \mathbf{0}\\ \frac{1}{\sigma}\boldsymbol{b}_K^{\mathsf{T}}\end{array}\right] = \mathcal{J}_K^{\mathsf{T}}.$$

Finally, Var $\left(\frac{\partial}{\partial\boldsymbol{\beta}}\text{Log}\left(\frac{1}{\sigma}\phi\left(\frac{\epsilon_t}{\sigma}\right)\right)\right) = \mathcal{I}_\beta = \left[\begin{array}{cc}\boldsymbol{C} & \mathbf{0}\\ \mathbf{0} & \frac{2}{\sigma^2}\end{array}\right]$, for some matrix $\boldsymbol{C}$ whose exact expression is not needed. Thus

$$\text{Var}(\boldsymbol{V}_t) = \left[\begin{array}{cc}\boldsymbol{I}_K & \mathcal{I}_K\\ \mathcal{I}_K^{\mathsf{T}} & \mathcal{I}_\beta\end{array}\right].$$

## APPENDIX C

We show that Var$(X_t) < \infty$. Without loss of generality, set $\sigma = 1$. This will be assumed here and in the next appendix. Since $Y_t$ is causal, we can write $Y_t = \sum_{j=0}^{\infty}\psi_j\epsilon_{t-j}$ and from (A.5)

$$\text{Var}(X_t) = E(w(\epsilon_t))^2 E\left(Y_{t-1} - \sum_{r=0}^{\infty}B_{r-1}\boldsymbol{\theta}^{\mathsf{T}}\boldsymbol{Y}_{t-1-r}^{(q)}\right)^2 = E(w(\epsilon_t))^2 E\left(\sum_{h=1}^{\infty}d_h\epsilon_{t-h}\right)^2$$

where $d_h = \psi_{h-1} - \sum_{\substack{r+j+l=h\\ 1\le j\le q\\ 0\le r,l\le h-1}}\psi_l\gamma_{r-1}\theta_j$. We now need the following lemma.

**Lemma 1.** *If the ARMA process (1) is causal and invertible, then* $\sum_{h=1}^{\infty}|d_h| < \infty$.

PROOF.

From (A.3), $\sum_{\substack{r+j+l=h\\ 1\le j\le q\\ 0\le r,l\le h-1}}\psi_l\gamma_{r-1}\theta_j = \sum_{k=1}^{h}\psi_{h-k}\left(\sum_{\substack{r+j=k\\ 1\le j\le q\\ 0\le r\le h-1}}\gamma_{r-1}\theta_j\right) = -\sum_{k=1}^{h}\psi_{h-k}\gamma_{k-1}$.

We have also $\sum_{h=1}^{\infty}\sum_{k=1}^{h}|\psi_{h-k}\gamma_{k-1}| = \sum_{h=0}^{\infty}\sum_{k=0}^{h}|\psi_{h-k}\gamma_k| = \sum_{k=0}^{\infty}|\gamma_k|\sum_{h=0}^{\infty}|\psi_h|$. Thus, $\sum_{h=1}^{\infty}|d_h| \le \sum_{h=1}^{\infty}|\psi_{h-1}| + \sum_{h=1}^{\infty}\sum_{k=1}^{h}|\psi_{h-k}\gamma_{k-1}| = \sum_{h=0}^{\infty}|\psi_h|\left(\sum_{k=0}^{\infty}|\gamma_k| + 1\right)$. But from Brockwell and Davis (1991), p.87, $\sum_{k=0}^{\infty}|\gamma_k|$ is finite. Since under the assumption of Theorem 1, $\sum_{j=0}^{\infty}|\psi_j| < \infty$ the lemma follows. $\square$

From this lemma, we conclude that $E\left[\sum_{h=1}^{\infty} d_h \epsilon_{t-h}\right]^2 = \sum_{h=1}^{\infty} d_h^2 < \infty$. Since

$$E(w(\epsilon_t))^2 = 4\int (L_k'(2\Phi(x)-1))^2 \phi^3(x)dx < \infty,$$

the result follows.

## APPENDIX D

Here we sketch the proof that the typical element $d_{4j}$ of inequality (A.6) vanishes. From $Y_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}$ and the fact that the remainder of a convergent series converges toward 0, we have

$$\begin{aligned}
\lim_{h\to\infty} d_{4j} = \lim_{h\to\infty} \sum_{r=0}^{\infty} |B_{r-1} E\left(w(\epsilon_t) Y_{t-1} Y_{t+h-r-j}\right)| \leq \\
\lim_{h\to\infty} |E(w(\epsilon_t))| \sum_{r=0}^{h-j} \left| B_{r-1} \sum_{a=0}^{\infty} \psi_a \psi_{a+h-j-r+1} \right| \\
\leq |E(w(\epsilon_t))| \lim_{h\to\infty} \left[ \sum_{a=0}^{m-1} |\psi_a| \sum_{r=0}^{a+h-j-m} |B_{r-1}\psi_{a+h-j-r+1}| + \sum_{r=a+h-j-m+1}^{h-j} |B_{r-1}\psi_{a+h-j-r+1}| \right. \\
\left. + \sum_{r=0}^{h-j} |B_{r-1}| \sum_{a=m}^{\infty} |\psi_a \psi_{a+h-j-r+1}| \right] \qquad \text{(D.1)}
\end{aligned}$$

where $m = \max\{p, q+1\} - p$. For the first term in the limit of (D.1), using the expression for $B_{r-1}$ in (A.4) and that of $\psi_{a+h-j-r+1}$ given in Brockwell and Davis (1991) eq. (3.3.6), we have

$$\begin{aligned}
&\sum_{r=q+1}^{a+h-j-m} |B_{r-1}\psi_{a+h-j-r+1}| = \\
&\sum_{r=q+1}^{a+h-j-m} \left| \sum_{b=1}^{k} \sum_{l=0}^{r_b-1} c_{bl} r^l \alpha_a^{-r} \sum_{b'=1}^{k'} \sum_{l'=0}^{r_{b'}-1} \alpha_{b'l'} (a+h-j-r+1)^{l'} \xi_{b'}^{-(a+h-j-r+1)} \right| \\
&\leq \sum_{b=1}^{k} \sum_{l=0}^{r_b-1} \sum_{b'=1}^{k'} \sum_{l'=0}^{r_{b'}-1} \sum_{d=0}^{l'} \binom{l'}{d} \left\{ |c_{bl}\alpha_{b'l'}| |\xi_{b'}^{-(a+h-j+1)} (a+h-j+1)^{l'-d}| \right. \\
&\left. \times \sum_{r=q+1}^{a+h-j-m} r^{l+d} |\alpha_a|^{-r} |\xi_{b'}|^r \right\}.
\end{aligned}$$

If $|\xi_{b'}| < |\alpha_a|$, the term in braces $\to 0$ as $h \to \infty$. Let $|\alpha_a| = 1+\epsilon_1 < |\xi_{b'}| = 1+\epsilon_2$ with $\epsilon_1, \epsilon_2 > 0$.

$$\left| \frac{(a+h-j+1)^{l'-d}}{\xi_{b'}^{(a+h-j+1)}} \right| \sum_{r=q+1}^{a+h-j-m} r^{l+d} |\alpha_a|^{-r} |\xi_{b'}|^r \leq \frac{|a+h-j+1|^{l'-d}}{|\xi_{b'}|^{a+h-j+1}} \sum_{r=0}^{a+h-j+1} r^{l+d} \left( \frac{|\xi_{b'}|}{|\alpha_a|} \right)^r . \qquad \text{(D.2)}$$

For all $\epsilon > 0$, there exist a $C, C'$ such that the left-hand side of (D.2) is bounded above by

$$C\frac{|a+h-j+1|^{l'-d}}{|\xi_{b'}|^{a+h-j+1}}\frac{\left(\frac{|\xi_{b'}|}{|\alpha_a|}+\epsilon\right)^{h+a-j+2}-1}{\frac{|\xi_{b'}|}{|\alpha_a|}+\epsilon-1} \le C'|a+h-j+1|^{l'-d}\left(\frac{\frac{|\xi_{b'}|}{|\alpha_a|}+\epsilon}{|\xi_{b'}|}\right)^{a+h-j+2}. \tag{D.3}$$

In (D.3), take $\epsilon > 0$ smaller than $\epsilon_1(1+\epsilon_2)/(1+\epsilon_1)$. Then the right hand side of (D.3) converges to 0 as $h \to \infty$. This shows that the first term in the limit of (D.1) converges to 0. It follows that the second term also converges toward 0. As for the last term in the limit, a similar argument yields that all terms on the right hand side of (D.1) converge to 0 so that $d_{4j} \to 0$.

## BIBLIOGRAPHY


[1] Anděl, J., 1997. On residual analysis for time series models. Kybernetika 33 (2), 161–170.

[2] Beiser, A., 1985. Distributions of $\sqrt{b_1}$ and $b_2$ for autoregressive errors. Ph.D. thesis, Boston University.

[3] Brockett, P. L., Hinich, M. J., Patterson, D., 1988. Bispectral-based tests for the detection of gaussianity and linearity in time series. J. Amer. Statist. Assoc. 83, 657–664.

[4] Brockwell, P. J., Davis, R. A., 1991. Time series: Theory and Methods, 2nd Edition. Springer-Verlag New York.

[5] Burn, D., 1987. Simulation of stationary time series. Proceedings of the 1987 Winter Simulation Conference , 289–294.

[6] Cromwell, J. B., Labys, W. C., Terraza, M., 1994. Univariate tests for time-series models. Sage Publications Inc, Thousand Oaks, California.

[7] D'Agostino, R. B., Stephens, M. A., 1986. Goodness-of-fit techniques. Statistics: TextBOOKs and Monographs, 68, New York: Marcel Dekker.

[8] Ducharme, G., Lafaye de Micheaux, P., 2002. Goodness-of-fit tests of normality for the innovations in arma models. Tech. rep., Technical report #02-02, Université Montpellier II.

[9] Epps, T. W., 1987. Testing that a stationnary time series is gaussian. Ann. Statist. 15 (4), 1683–1698.

[10] Frances, P., 1998. Time series models for business and economic forecasting. Cambridge University Press, Cambridge.

[11] Gasser, T., 1975. Goodness-of-fit tests for correlated data. Biometrika 62, 563–570.

[12] Gouriéroux, C., Monfort, A., 1995. Séries temporelles et modèles dynamiques, 2nd Edition. Economica.

[13] Granger, C. W. J., 1976. Tendency towards normality of linear combinations of random variables. Metrika 23 (4), 237–248.

[14] Henze, N., 1997. Do components of smooth tests of fit have diagnostic properties? Metrika 45, 121–130.

[15] Heuts, R., Rens, S., 1986. Testing normality when observations satisfy a certain low order arma-scheme. Computat. Statist. Quarterly 1, 49–60.

[16] Hinich, M. J., 1982. Testing for gaussianity and linearity of a stationary time series. J. Time Ser. Anal. 3 (3), 169–176.

[17] Hipel, K. W., McLeod, A. I., 1994. Time series modelling of water resources and environmental systems. [Elsevier Science Publishing Co., New York; North-Holland Publishing Co., Amsterdam] (New York; Amsterdam).

[18] Janic-Wroblewska, A., Ledwina, T., 2000. Data driven rank test for two-sample problem. Scand. J. Statist. 27, 281–298.

[19] Jarque, C., Bera, A., 1987. A test for normality of observations and regression residuals. Internat. Statist. Review 55 (2), 163–172.

[20] Kallenberg, W., Ledwina, T., 1997a. Data driven smooth tests for composite hypotheses: comparison of powers. J. Statist. Comput. Simul. 59 (2), 101–121.

[21] Kallenberg, W., Ledwina, T., 1997b. Data-driven smooth tests when the hypothesis is composite. J. Amer. Statist. Assoc. 92 (439), 1094–1104.

[22] Koul, H. L., Stute, W., 1999. Nonparametric model checks for time series. Ann. Statist. 27 (1), 204–236.

[23] Ledwina, T., 1994. Data-driven version of Neyman's smooth test of fit. J. Amer. Statist. Assoc. 89 (427), 1000–1005.

[24] Lee, S., Na, S., 2002. On the Bickel-Rosenblatt test for first-order autoregressive models. Statist. Probab. Lett. 56 (1), 23–35.

[25] Lomnicki, Z., 1961. Tests for departure from normality in the case of linear stochastic processes. Metrika 4, 37–62.

[26] Lutkepohl, H., Schneider, W., 1989. Testing for normality of autoregressive time series. Comput. Statist. Quaterly 2, 151–168.

[27] Moore, D. S., 1982. The effect of dependence on chi squared tests of fit. Ann. Statist. 10 (4), 1163–1171.

[28] Neyman, J., 1937. Smooth test for goodness of fit. Skand. Aktuar. 20, 149–199.

[29] Ojeda, R., Cardoso, J., Moulines, E., 1997. Asymptotically invariant gaussianity test for causal invertible time series. Proc. of IEEE international conference on Acoustics, Speech and Signal Processing 5, 3713–3716.

[30] Pierce, D. A., Gray, R. J., 1985. Goodness-of-fit tests for censored survival data. Ann. Statist. 13 (2), 552–563.

[31] Rayner, J., Best, D., 1989. Smooth Tests of Goodness-of-Fit. Oxford: Oxford University Press.

[32] Sansone, G., 1959. Orthogonal functions. New York: Interscience.

[33] Schwarz, G., 1978. Estimating the dimension of a model. Ann. Statist. 6 (2), 461–464.

[34] Shea, B. L., 1987. Estimation of multivariate time series. J. Time Ser. Anal. 8 (1), 95–109.

[35] Weisberg, S., Bingham, C., 1975. An approximate analysis of variance test for non-normality suitable for machine calculation. Technometrics 17, 133–134.

TABLE 4.1. Distribution (in % of the number of parameter points) of the empirical $p$-values (based on 10000 replications) for the tests based on $\mathcal{R}_3$ and $\mathcal{R}_{\hat{K}}(2)$ among 5 sub-intervals.

| $\mathcal{R}_3$ | | | Observed level | | | | | Min |
|---|---|---|---|---|---|---|---|---|
| Model | T | $\alpha$ | $I_1$ | $I_2$ | $I_3$ | $I_4$ | $I_5$ | $p$-level |
| | 50 | 5% | 18.8 | 68.8 | 12.5 | 0 | 0 | 2.76 |
| | 100 | 5% | 1.6 | 50.0 | 48.4 | 0 | 0 | 3.41 |
| MA(2) | 200 | 5% | 0 | 9.4 | 89.1 | 1.6 | 0 | 4.04 |
| | 50 | 10% | 23.4 | 53.1 | 23.4 | 0 | 0 | 6.49 |
| (64 points) | 100 | 10% | 6.3 | 20.3 | 73.4 | 0 | 0 | 7.96 |
| | 200 | 10% | 0 | 7.8 | 90.6 | 1.6 | 0 | 8.92 |
| | 50 | 5% | 47.2 | 46.4 | 6.4 | 0 | 0 | 2.43 |
| | 100 | 5% | 8.0 | 71.6 | 20.4 | 0 | 0 | 2.98 |
| ARMA(1,2) | 200 | 5% | 0.8 | 32.4 | 66.4 | 0.4 | 0 | 3.32 |
| | 50 | 10% | 65.6 | 24.0 | 10.4 | 0 | 0 | 6.20 |
| (250 points) | 100 | 10% | 21.6 | 35.2 | 42.8 | 0.4 | 0 | 6.80 |
| | 200 | 10% | 4.0 | 19.6 | 75.6 | 0.8 | 0 | 7.42 |
| | 50 | 5% | 41.2 | 57.1 | 1.7 | 0 | 0 | 2.56 |
| | 100 | 5% | 5.1 | 74.1 | 20.8 | 0 | 0 | 3.09 |
| ARMA(2,2) | 200 | 5% | 0.3 | 27.9 | 71.8 | 0 | 0 | 3.47 |
| | 50 | 10% | 57.8 | 37.4 | 4.8 | 0 | 0 | 6.24 |
| (294 points) | 100 | 10% | 21.1 | 33.7 | 45.2 | 0 | 0 | 6.88 |
| | 200 | 10% | 3.1 | 18.0 | 78.6 | 0.3 | 0 | 7.86 |
| $\mathcal{R}_{\hat{K}}(2)$ | | | Observed level | | | | | Min |
| Model | T | $\alpha$ | $I_1$ | $I_2$ | $I_3$ | $I_4$ | $I_5$ | $p$-level |
| | 50 | 5% | 0 | 9.4 | 46.9 | 43.7 | 0 | 4.12 |
| | 100 | 5% | 0 | 14.1 | 68.8 | 17.2 | 0 | 4.17 |
| MA(2) | 200 | 5% | 0 | 7.8 | 87.5 | 4.7 | 0 | 4.23 |
| | 50 | 10% | 6.3 | 14.1 | 62.5 | 17.2 | 0 | 7.78 |
| (64 points) | 100 | 10% | 6.3 | 6.3 | 81.3 | 6.3 | 0 | 8.19 |
| | 200 | 10% | 0 | 6.3 | 89.1 | 4.7 | 0 | 8.83 |
| | 50 | 5% | 0 | 38.8 | 46.8 | 14.4 | 0 | 3.53 |
| | 100 | 5% | 0 | 34.8 | 59.2 | 6.0 | 0 | 3.74 |
| ARMA(1,2) | 200 | 5% | 0 | 23.2 | 75.2 | 1.6 | 0 | 3.80 |
| | 50 | 10% | 24.8 | 31.6 | 35.2 | 8.4 | 0 | 7.06 |
| (250 points) | 100 | 10% | 13.2 | 27.2 | 57.6 | 2.0 | 0 | 7.33 |
| | 200 | 10% | 4.4 | 18.0 | 76.0 | 1.6 | 0 | 7.61 |
| | 50 | 5% | 0 | 32.0 | 55.4 | 12.6 | 0 | 3.65 |
| | 100 | 5% | 0 | 31.0 | 62.9 | 6.1 | 0 | 3.75 |
| ARMA(2,2) | 200 | 5% | 0 | 23.8 | 75.9 | 0.3 | 0 | 3.89 |
| | 50 | 10% | 21.4 | 30.3 | 47.0 | 1.4 | 0 | 7.14 |
| (294 points) | 100 | 10% | 11.2 | 24.8 | 62.6 | 1.4 | 0 | 7.51 |
| | 200 | 10% | 2.7 | 16.7 | 80.6 | 0 | 0 | 7.62 |

TABLE 4.2. Distribution (in % of the number of parameter points) of the empirical $p$-values (based on 10000 replications) of various tests for the ARMA(1,2) model. $\mathcal{AD}$=Anderson-Darling, $\mathcal{WB}$=Weisberg-Bingham, $\mathcal{JB}$=Jarque-Bera and $\mathcal{R}_{\hat{K}}(1) = \mathcal{R}_{\hat{K}}$ with $d = 1$.

| Test | | | Observed level | | | | | Min |
|---|---|---|---|---|---|---|---|---|
| Model | T | $\alpha$ | $I_1$ | $I_2$ | $I_3$ | $I_4$ | $I_5$ | $p$-level |
| $\mathcal{AD}$ | 50 | 5% | 43.8 | 24.4 | 20.0 | 6.8 | 0 | 0.54 |
| | 100 | 5% | 32.0 | 34.8 | 33.2 | 0 | 0 | 0.92 |
| | 200 | 5% | 11.6 | 38.0 | 49.6 | 0.8 | 0 | 1.50 |
| | 50 | 10% | 41.2 | 16.0 | 22.8 | 18.4 | 1.6 | 3.38 |
| | 100 | 10% | 23.2 | 24.4 | 48.0 | 3.6 | 0.8 | 3.93 |
| | 200 | 10% | 9.6 | 13.2 | 70.0 | 6.8 | 0.4 | 4.65 |
| $\mathcal{WB}$ | 50 | 5% | 61.2 | 23.6 | 15.2 | 0 | 0 | 0.57 |
| | 100 | 5% | 39.2 | 46.0 | 14.4 | 0.4 | 0 | 0.93 |
| | 200 | 5% | 10.8 | 30.4 | 56.0 | 2.8 | 0 | 1.60 |
| | 50 | 10% | 56.8 | 16.4 | 25.6 | 1.2 | 0 | 2.96 |
| | 100 | 10% | 37.2 | 47.2 | 15.6 | 0 | 0 | 3.55 |
| | 200 | 10% | 15.6 | 36.0 | 46.0 | 2.4 | 0 | 4.71 |
| $\mathcal{JB}$ | 50 | 5% | 71.2 | 28.8 | 0 | 0 | 0 | 3.13 |
| | 100 | 5% | 0.4 | 99.2 | 0.4 | 0 | 0 | 3.13 |
| | 200 | 5% | 0 | 85.2 | 14.4 | 0.4 | 0 | 4.18 |
| | 50 | 10% | 100 | 0 | 0 | 0 | 0 | 4.96 |
| | 100 | 10% | 100 | 0 | 0 | 0 | 0 | 5.88 |
| | 200 | 10% | 98.8 | 1.2 | 0 | 0 | 0 | 7.23 |
| $\mathcal{R}_{\hat{K}}(1)$ | 50 | 5% | 0 | 58.0 | 27.6 | 14.4 | 0 | 3.52 |
| | 100 | 5% | 10.0 | 48.8 | 39.2 | 2.0 | 0 | 3.32 |
| | 200 | 5% | 8.4 | 33.6 | 56.4 | 1.6 | 0 | 3.39 |
| | 50 | 10% | 43.6 | 20.8 | 26.0 | 9.6 | 0 | 4.69 |
| | 100 | 10% | 26.0 | 24.8 | 48.0 | 1.2 | 0 | 4.81 |
| | 200 | 10% | 8.8 | 16.4 | 70.4 | 4.4 | 0 | 5.79 |

TABLE 4.3. Empirical power (based on 10000 replications with $\alpha = 10\%$) of various tests when $T = 50$. The part above the line in the middle of the table corresponds to symmetric alternatives while those below are skewed. The distributions are ordered according to increasing kurtosis. The ARMA(2,1) model has parameter $(\boldsymbol{\varphi}, \theta_1) = (-0.8, -0.1, 0.7)$, the ARMA(1,2) model has parameter $(\varphi_1, \boldsymbol{\theta}) = (-0.7, 0.4, 0.5)$ while the ARMA(2,2) model has parameter $(\boldsymbol{\varphi}, \boldsymbol{\theta}) = (-1.05, -0.4, 0.15, 0.85)$.

| $T = 50$ | ARMA(2,1) | | | ARMA(1,2) | | | ARMA(2,2) | | |
|---|---|---|---|---|---|---|---|---|---|
| Alternatives | $\mathcal{R}_3$ | $\mathcal{R}_{\hat{K}}(2)$ | $\mathcal{WB}$ | $\mathcal{R}_3$ | $\mathcal{R}_{\hat{K}}(2)$ | $\mathcal{WB}$ | $\mathcal{R}_3$ | $\mathcal{R}_{\hat{K}}(2)$ | $\mathcal{WB}$ |
| SB(0;0.5) | 83.19 | 84.76 | 28.93 | 73.16 | 74.90 | 22.47 | 63.85 | 65.44 | 17.57 |
| TU(1.5) | 66.94 | 67.96 | 18.27 | 57.86 | 59.43 | 15.07 | 49.17 | 50.50 | 13.62 |
| TU(0.7) | 44.47 | 45.64 | 12.42 | 38.69 | 39.44 | 11.02 | 32.58 | 33.88 | 10.76 |
| Logistic(1) | 20.74 | 22.75 | 19.02 | 18.97 | 20.87 | 17.10 | 18.59 | 20.36 | 17.60 |
| TU(10) | 94.64 | 96.64 | 83.60 | 89.57 | 91.62 | 74.31 | 85.14 | 87.26 | 65.78 |
| SC(0.05;3) | 33.65 | 37.38 | 35.98 | 32.82 | 35.65 | 34.40 | 30.81 | 34.69 | 32.37 |
| SC(0.2;5) | 96.36 | 96.77 | 92.84 | 94.21 | 94.72 | 89.12 | 92.28 | 92.96 | 85.62 |
| SC(0.05;5) | 62.33 | 65.22 | 63.63 | 61.43 | 63.86 | 61.81 | 58.90 | 62.20 | 59.77 |
| SC(0.05;7) | 74.05 | 76.12 | 75.32 | 73.00 | 75.22 | 73.89 | 72.28 | 74.04 | 72.50 |
| SU(0;1) | 75.96 | 76.49 | 66.57 | 71.73 | 72.44 | 62.20 | 68.60 | 69.79 | 59.99 |
| SB(0.533;0.5) | 91.09 | 89.76 | 59.41 | 83.62 | 82.09 | 48.17 | 76.29 | 74.04 | 36.87 |
| SB(1;1) | 53.75 | 56.94 | 32.60 | 45.94 | 48.41 | 26.18 | 42.03 | 44.17 | 23.84 |
| LC(0.2;3) | 55.58 | 57.79 | 29.72 | 49.58 | 52.11 | 26.19 | 43.88 | 46.16 | 22.62 |
| Weibull(2) | 28.10 | 30.72 | 21.25 | 25.63 | 28.66 | 18.52 | 24.32 | 26.85 | 18.68 |
| LC(0.1;3) | 44.10 | 43.85 | 35.21 | 40.62 | 40.54 | 31.38 | 37.00 | 37.25 | 27.97 |
| $\chi^2$ (df.=10) | 41.41 | 45.84 | 34.72 | 37.80 | 41.98 | 31.09 | 35.04 | 38.91 | 29.66 |
| LC(0.05;3) | 29.50 | 31.25 | 28.28 | 28.05 | 29.86 | 26.18 | 25.91 | 28.32 | 24.17 |
| LC(0.1;5) | 96.10 | 96.00 | 95.07 | 93.40 | 93.03 | 90.44 | 91.04 | 89.90 | 86.06 |
| SU(-1;2) | 37.88 | 38.92 | 33.93 | 34.38 | 36.29 | 31.12 | 33.54 | 65.40 | 29.87 |
| $\chi^2$ (df.=4) | 76.13 | 80.54 | 69.78 | 71.19 | 75.76 | 63.30 | 65.96 | 70.30 | 57.70 |
| LC(0.05;5) | 81.48 | 83.98 | 84.41 | 78.07 | 80.38 | 80.80 | 75.48 | 77.71 | 77.58 |
| LC(0.05;7) | 94.26 | 94.74 | 96.28 | 94.05 | 94.73 | 96.39 | 93.44 | 94.24 | 95.88 |
| SU(1;1) | 96.26 | 96.15 | 93.98 | 94.18 | 94.17 | 91.39 | 93.14 | 93.17 | 90.14 |
| LN(0;1) | 99.52 | 99.68 | 99.24 | 98.74 | 98.85 | 98.02 | 97.89 | 98.16 | 96.83 |

# CHAPITRE 3

# A multivariate empirical characteristic function test of independence with normal marginals

Cet article a été soumis à la revue *Journal of Multivariate Analysis*.

Comme la coutume dans cette discipline le veut, l'ordre alphabétique des auteurs a été respecté.

Voici la liste des contributions principales de Pierre Lafaye de Micheaux à cet article :

- Conduite d'une importante recherche bibliographique.
- Démonstration des Théorèmes 2.1, 2.3, 3.1, 3.2 et 3.3.
- Démonstration d'une partie de la preuve du Théorème 2.2.
- Démonstration de l'essentiel de la preuve du Théorème 2.4.
- Conduite générale de la recherche pour l'intégralité de la section 4.
- Conception, écriture et validation des programmes C++ et simulations.
- Rédaction d'une bonne partie de l'article en anglais.

# A multivariate empirical characteristic function test of independence with normal marginals

Martin Bilodeau and Pierre Lafaye de Micheaux

*Département de mathématiques et de statistique, Université de Montréal, Canada*

**Abstract**
This paper proposes a semi-parametric test of independence (or serial independence) between marginal vectors each of which is normally distributed but without assuming the joint normality of these marginal vectors. The test statistic is a Cramér-von Mises functional of a process defined from the empirical characteristic function. This process is defined similarly as the process of Ghoudi et al. (2001) built from the empirical distribution function and used to test for independence between univariate marginal variables. The test statistic can be represented as a V statistic. It is consistent to detect any form of dependence. The weak convergence of the process is derived. The asymptotic distribution of the Cramér-von Mises functionals are approximated by the Cornish-Fisher expansion using a recursive formula for cumulants and by the numerical evaluations of the eigenvalues in the inversion formula. The test statistic is finally compared with Wilks' statistic for testing the parametric hypothesis of independence in the one-way MANOVA model with random effects.

*Key words:* Characteristic function, Independence, Multivariate Analysis, Serial independence, Stochastic processes
*1991 MSC:* 62H15, 62M99

## 1. INTRODUCTION

Different characterizations have led to various tests of independence. Let $p \geq 1$ be a fixed integer. Consider a partitioned random vector $\boldsymbol{\epsilon} = \left(\boldsymbol{\epsilon}^{(1)}, \ldots, \boldsymbol{\epsilon}^{(p)}\right)$ made up of $p$ $q$-dimensional subvectors and a corresponding partitioned $\boldsymbol{t} = \left(\boldsymbol{t}^{(1)}, \ldots, \boldsymbol{t}^{(p)}\right)$, for any fixed vector $\boldsymbol{t}$. Independence of the subvectors may be characterized with the joint distribution function or characteristic function as

$$K_p(\boldsymbol{t}) = \prod_{k=1}^{p} K^{(k)}(\boldsymbol{t}^{(k)}), \quad (1.1)$$

$$C_p(\boldsymbol{t}) = \prod_{k=1}^{p} C^{(k)}(\boldsymbol{t}^{(k)}), \quad (1.2)$$

where $K_p$ and $C_p$ are, respectively, the joint distribution function and joint characteristic function. The marginal versions are $K^{(k)}$ and $C^{(k)}$ for $k = 1, \ldots, p$. In the univariate setting ($q = 1$) Blum et al. (1961) proposed an empirical process based on

(1.1), whereas Csörgő (1981a) defined a similar process based on (1.2). Feuerverger (1993) proposes an empirical characteristic function version of the Blum et al. (1961) test statistic. He points out difficulties with dimensions above 2.

Recently, in the univariate setting, Ghoudi et al. (2001) introduced a new process based on their novel characterization of independence which is now presented. This characterization for $p = 3$ is implicit in the paper of Blum et al. (1961). For any $A \subset I_p = \{1, \dots, p\}$ and any $\boldsymbol{t} \in \mathbb{R}^p$, let

$$\mu_A(\boldsymbol{t}) = \sum_{B \subset A} (-1)^{|A \backslash B|} K_p(\boldsymbol{t}^B) \prod_{j \in A \backslash B} K^{(j)}(\boldsymbol{t}^{(j)}).$$

The notation $|A|$ stands for cardinality of the set $A$ and the convention $\prod_{\emptyset} = 1$ is adopted. The vector $\boldsymbol{t}^B$ is used to make a selection of components of $\boldsymbol{t}$ according to the set $B$,

$$(\boldsymbol{t}^B)^{(i)} = \begin{cases} \boldsymbol{t}^{(i)}, & i \in B; \\ \infty, & i \in I_p \backslash B. \end{cases}$$

Independence can be characterized: $\boldsymbol{\epsilon}^{(1)}, \dots, \boldsymbol{\epsilon}^{(p)}$ are independent if and only if $\mu_A \equiv 0$, for all $A \subset I_p$ satisfying $|A| > 1$. Cramér-von Mises type functional of an associated process then leads them to a non-parametric test of independence in the non-serial or serial situation. The interest in their process resides in the simple form of the covariance which is expressed as a product of covariance functions of Brownian bridge.

In the multivariate setting ($q \geq 1$), the present paper proposes tests of independence, when subvectors or marginals are normally distributed, built from a process relying on a similar independence characterization based on characteristic functions. Note that the subvectors are not assumed to be jointly multinormal in which case independence can be tested parametrically with covariances using likelihood ratio tests. Namely, the marginals $\boldsymbol{\epsilon}^{(1)}, \dots, \boldsymbol{\epsilon}^{(p)}$ are independent if and only if $\mu_A \equiv 0$, for all $A \subset I_p$, $|A| > 1$. Here,

$$\mu_A(\boldsymbol{t}) = \sum_{B \subset A} (-1)^{|A \backslash B|} C_p(\boldsymbol{t}^B) \prod_{j \in A \backslash B} C^{(j)}(\boldsymbol{t}^{(j)}),$$

where

$$(\boldsymbol{t}^B)^{(i)} = \begin{cases} \boldsymbol{t}^{(i)}, & i \in B; \\ \mathbf{0}, & i \in I_p \backslash B. \end{cases}$$

Note that the subvectors are not assumed to be jointly multinormal in which case independence can be tested parametrically with covariances using likelihood ratio tests. Normality of the marginals will often be approximately satisfied. For example, data analysis of regression models on which Box-Cox transformations are done in the first stage are common in practice. One should bear in mind, however, that the asymptotic of the tests of independence proposed here would not consider the Box-Cox transformation as data dependent but as a fixed transformation. Goodness-of-fit tests of normality after a data dependent Box-Cox transformation by Chen et al. (2002) is a result in this direction; it will not be pursued here for tests of independence.

The non-serial and serial problems are considered. It is shown that the asymptotic distribution of the proposed process is the same in both cases under the null hypothesis of independence. Moreover, it is established that the estimation of the unknown mean vector and positive definite covariance matrix of the normal marginals does not affect the asymptotic distribution of the process. The proposed Cramér-von Mises type of test statistic is related to V-statistics for which de Wet and Randles (1987) studied the effect of estimating unknown parameters. Norm on Euclidian spaces $\mathbb{R}^m$ will be denoted $\|\cdot\|$, whereas $|\cdot|$ will be the norm on the complex field $\mathbb{C}$. We also let $C(\mathbb{R}^{pq}, \mathbb{C})$ be the space of continuous functions from $\mathbb{R}^{pq}$ to $\mathbb{C}$.

# 2. TESTING INDEPENDENCE: THE NON-SERIAL SITUATION

## 2.1. The case of known parameters

Let $\boldsymbol{\epsilon} = \left(\boldsymbol{\epsilon}^{(1)}, \ldots, \boldsymbol{\epsilon}^{(p)}\right) \in \mathbb{R}^{pq}$ denote a partition into $p$ subvectors and $\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_n$ be an i.i.d. sample of such $(pq)$-dimensional random vectors. Suppose that the subvectors of the random vectors $\boldsymbol{\epsilon}_i$ all have the same $N_q(\mathbf{0}, \boldsymbol{I})$ normal distribution, with characteristic function $\phi$. The problem is that of testing the independence of the marginals, that is the independence of $\boldsymbol{\epsilon}^{(1)}, \ldots, \boldsymbol{\epsilon}^{(p)}$. This non-serial problem with known parameters is of very limited practical importance. However, it serves as a prototype on which subsequent results are based.

Following Ghoudi et al. (2001), for any $A \subset I_p = \{1, \ldots, p\}$ and any $\boldsymbol{t} = (\boldsymbol{t}^{(1)}, \ldots, \boldsymbol{t}^{(p)}) \in \mathbb{R}^{pq}$, let

$$R_{n,A}(\boldsymbol{t}) = \sqrt{n} \sum_{B \subset A} (-1)^{|A \setminus B|} \phi_{n,p}(\boldsymbol{t}^B) \prod_{i \in A \setminus B} \phi(\boldsymbol{t}^{(i)}) \tag{2.1}$$

where

$$\phi_{n,p}(\boldsymbol{t}) = \frac{1}{n} \sum_{j=1}^{n} \exp(\mathrm{i}\langle \boldsymbol{t}, \boldsymbol{\epsilon}_j \rangle) \tag{2.2}$$

is the empirical characteristic function of the sample. The notation $\langle \cdot, \cdot \rangle$ is for the usual inner product between vectors.

The asymptotic behaviour of these processes is stated next. It states that for different subsets $A$ the associated processes are asymptotically independent, each process being asymptotically Gaussian with a covariance function of a particularly simple form. Specifically, the covariance function is a product of covariance functions of the type encountered by Feuerverger and Mureika (1977), Csörgő (1981a) or Marcus (1981), for the empirical characteristic function process. Another process defined by Csörgő (1985) has a covariance of a more complicated structure.

**Theorem 2.1.** *If $\boldsymbol{\epsilon}_1^{(1)}, \ldots, \boldsymbol{\epsilon}_1^{(p)}$ are independent, the collection of processes $\{R_{n,A} : |A| > 1\}$ converge in $C(\mathbb{R}^{pq}, \mathbb{C})$ to independent zero mean complex Gaussian processes $\{R_A : |A| > 1\}$ having covariance function given by*

$$C_A(\boldsymbol{s}, \boldsymbol{t}) = \prod_{k \in A} [\phi(\boldsymbol{t}^{(k)} - \boldsymbol{s}^{(k)}) - \phi(\boldsymbol{t}^{(k)})\phi(\boldsymbol{s}^{(k)})] \tag{2.3}$$

*and pseudo-covariance function given by*

$$\overline{C}_A(\boldsymbol{s}, \boldsymbol{t}) = E[R_A(\boldsymbol{s}) R_A(\boldsymbol{t})] = C_A(-\boldsymbol{s}, \boldsymbol{t}). \tag{2.4}$$

The multinomial formula (Ghoudi et al. (2001)) yields the equivalent representation

$$R_{n,A}(\boldsymbol{t}) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \prod_{k \in A} [\exp(\mathrm{i}\langle \boldsymbol{t}^{(k)}, \boldsymbol{\epsilon}_j^{(k)} \rangle) - \phi(\boldsymbol{t}^{(k)})]. \tag{2.5}$$

This i.i.d. average representation is used in the proof of Theorem 2.1.

### 2.2. The case of unknown parameters

The context is the same as in the preceeding subsection except that the components of the random vectors $\boldsymbol{\epsilon}_i$ in the sample now have all the same $N_q(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ normal distribution, where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, positive definite, are unknown. The problem again is that of testing for independence of the marginals, that is the independence of the components $\boldsymbol{\epsilon}^{(1)}, \ldots, \boldsymbol{\epsilon}^{(p)}$.

First, define the standardized residual vectors $\boldsymbol{e}_j^{(k)} = \boldsymbol{S}^{-\frac{1}{2}}(\boldsymbol{\epsilon}_j^{(k)} - \overline{\boldsymbol{\epsilon}})$, where $\boldsymbol{S} = \frac{1}{np} \sum_{j=1}^{n} \sum_{k=1}^{p} (\boldsymbol{\epsilon}_j^{(k)} - \overline{\boldsymbol{\epsilon}})(\boldsymbol{\epsilon}_j^{(k)} - \overline{\boldsymbol{\epsilon}})^{\mathsf{T}}$ and $\overline{\boldsymbol{\epsilon}} = \frac{1}{np} \sum_{j=1}^{n} \sum_{k=1}^{p} \boldsymbol{\epsilon}_j^{(k)}$ are, respectively, the sample covariance matrix and the sample mean. Also, let $\overline{\boldsymbol{\epsilon}}^{(k)} = \frac{1}{n} \sum_{j=1}^{n} \boldsymbol{\epsilon}_j^{(k)}$ be the sample mean of the $k$th subvectors.

The underlying process is the same apart from the unknown parameters which are replaced by their sample estimates. The plug-in process is thus

$$\hat{R}_{n,A}(\boldsymbol{t}) = \sqrt{n} \sum_{B \subset A} (-1)^{|A \setminus B|} \hat{\phi}_{n,p}(\boldsymbol{t}^B) \prod_{i \in A \setminus B} \phi(\boldsymbol{t}^{(i)}) \tag{2.6}$$

where

$$\hat{\phi}_{n,p}(\boldsymbol{t}) = \frac{1}{n} \sum_{j=1}^{n} \exp(\mathrm{i}\langle \boldsymbol{t}, \boldsymbol{e}_j \rangle) \tag{2.7}$$

is the empirical characteristic function based on standardized residuals $\boldsymbol{e}_j = \left(\boldsymbol{e}_j^{(1)}, \ldots, \boldsymbol{e}_j^{(p)}\right) \in \mathbb{R}^{pq}$. The asymptotic behaviour of these processes is stated next, the main conclusion being that the estimation of the unknown parameters does not affect the asymptotic distribution.

**Theorem 2.2.** *If $\boldsymbol{\epsilon}_1^{(1)}, \ldots, \boldsymbol{\epsilon}_1^{(p)}$ are independent, the processes $\{\hat{R}_{n,A} : |A| > 1\}$ converge in $C(\mathbb{R}^{pq}, \mathbb{C})$ to independent zero mean complex Gaussian processes $\{R_A : |A| > 1\}$ having covariance and pseudo-covariance functions respectively given by $C_A(\boldsymbol{s}, \boldsymbol{t})$ and $\overline{C}_A(\boldsymbol{s}, \boldsymbol{t})$ in (2.3) and (2.4).*

The same multinomial formula (Ghoudi et al. (2001)) yields the representation

$$\hat{R}_{n,A}(\boldsymbol{t}) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \prod_{k \in A} [\exp(\mathrm{i}\langle \boldsymbol{t}^{(k)}, \boldsymbol{e}_j^{(k)} \rangle) - \phi(\boldsymbol{t}^{(k)})]. \tag{2.8}$$

The Cramér-von Mises test statistic proposed is $nT_{n,b,A}$, where for a given subset $A$

$$T_{n,b,A} = \frac{1}{n} \int |\hat{R}_{n,A}(\boldsymbol{t})|^2 \varphi_b(\boldsymbol{t}) \mathrm{d}\boldsymbol{t}, \tag{2.9}$$

where $\varphi_b$ is the $N_{pq}(\mathbf{0}, b^2\boldsymbol{I})$ density which acts as a weighting function. The multinomial representation and this appropriate weighting allow this test statistic to be computed explicitely as

$$\frac{1}{n^2}\sum_{l=1}^{n}\sum_{l'=1}^{n}\prod_{k\in A}\Bigg\{ \quad \exp \quad \left[-\frac{b^2}{2}\|\boldsymbol{e}_l^{(k)} - \boldsymbol{e}_{l'}^{(k)}\|^2\right] \tag{2.10}$$

$$-(b^2+1)^{-\frac{q}{2}} \quad \exp \quad \left[-\frac{1}{2}\frac{b^2}{b^2+1}\|\boldsymbol{e}_l^{(k)}\|^2\right] \tag{2.11}$$

$$-(b^2+1)^{-\frac{q}{2}} \quad \exp \quad \left[-\frac{1}{2}\frac{b^2}{b^2+1}\|\boldsymbol{e}_{l'}^{(k)}\|^2\right] + (2b^2+1)^{-\frac{q}{2}}\Bigg\}. \tag{2.12}$$

Since squared Mahalanobis type statistics are affine invariant it follows that $T_{n,b,A}$ is affine invariant. Thus, the asymptotic distribution of this statistic does not depend on unknown parameters.

It should be noted that the functional (2.9) defining this test statistic is not continuous; it is not even defined on $C(\mathbb{R}^{pq}, \mathbb{C})$ but only on the subset of squared-integrable functions with respect to the measure $\varphi_b(\boldsymbol{t})d\boldsymbol{t}$. Thus, the continuity theorem as in Billingsley (1968) can not be invoked. In order to obtain the asymptotic distribution of this functional, the following generalization of Theorem 3.3 of Kellermeier (1980) on a uniform integrability condition is proposed. Let $\mathbb{R}_j^{pq}$ be the ball of radius $j$ centered at zero in $\mathbb{R}^{pq}$.

**Theorem 2.3.** *Let $\boldsymbol{y}_n$ and $\boldsymbol{y}$ be random elements of $C(\mathbb{R}^{pq}, \mathbb{C})^2$ such that $\boldsymbol{y}_n \xrightarrow{\mathcal{D}} \boldsymbol{y}$ on all compact balls. Let $f : \mathbb{C}^2 \to \mathbb{R}$ be a continuous function and let G be a probability measure on $\mathbb{R}^{pq}$. Define $w_n = \int f(\boldsymbol{y}_n(\boldsymbol{t}))dG(\boldsymbol{t})$ and $w = \int f(\boldsymbol{y}(\boldsymbol{t}))dG(\boldsymbol{t})$. Suppose that $w_n$ and $w$ are well defined with probability one. Moreover, suppose that there exists $\alpha \geq 1$ such that for all $\epsilon > 0$,*

$$\lim_{j\to\infty}\limsup_{n\to\infty}\int_{\mathbb{R}^{pq}\setminus\mathbb{R}_j^{pq}} E|f(\boldsymbol{y}_n(\boldsymbol{t}))|^\alpha dG(\boldsymbol{t}) = 0. \tag{2.13}$$

*Then $w_n \xrightarrow{\mathcal{D}} w$ as $n \to \infty$.*

Using Theorem 2.3, the joint convergence of the Cramér-von Mises functionals can be established.

**Theorem 2.4.**

$$\int \left(|\hat{R}_{n,A}(\boldsymbol{t})|^2, |\hat{R}_{n,B}(\boldsymbol{t})|^2\right)\varphi_b(\boldsymbol{t})d\boldsymbol{t} \xrightarrow{\mathcal{D}} \int \left(|R_A(\boldsymbol{t})|^2, |R_B(\boldsymbol{t})|^2\right)\varphi_b(\boldsymbol{t})d\boldsymbol{t},$$

*where integrals are computed componentwise.*

All possible subsets $A$ can then be simultaneously accounted for by combining the test statistics as in

$$S_n = n\sum_{|A|>1} T_{n,b,A} \tag{2.14}$$

or

$$M_n = n \max_{|A|>1} T_{n,b,A}. \tag{2.15}$$

### 2.3. Relation to V-statistics

The statistic $T_{n,b,A}$ is in fact a V-statistic as in de Wet and Randles (1987). It can be represented as

$$T_{n,b,A} = \frac{1}{n^2} \sum_{l=1}^{n} \sum_{l'=1}^{n} h(\boldsymbol{\epsilon}_l, \boldsymbol{\epsilon}_{l'}; \hat{\boldsymbol{\lambda}}_n), \tag{2.16}$$

where $\hat{\boldsymbol{\lambda}}_n = (\bar{\boldsymbol{\epsilon}}, \boldsymbol{S})$ consistently estimates the true parameter $\boldsymbol{\lambda} = (\mathbf{0}, \boldsymbol{I})$. The function $h$ at an arbitrary $\boldsymbol{\gamma} = (\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is defined as

$$h(\boldsymbol{\epsilon}_l, \boldsymbol{\epsilon}_{l'}; \boldsymbol{\gamma}) = \int g(\boldsymbol{\epsilon}_l, \boldsymbol{t}; \boldsymbol{\gamma}) g(\boldsymbol{\epsilon}_{l'}, \boldsymbol{t}; \boldsymbol{\gamma}) \varphi_b(\boldsymbol{t}) \mathrm{d}\boldsymbol{t},$$

where, from elementary properties of integrals of odd functions, the function $g$ can be taken real-valued

$$\begin{aligned} g(\boldsymbol{\epsilon}_l, \boldsymbol{t}; \boldsymbol{\gamma}) \;=\; \prod_{k \in A} \Big[ & \cos(\langle \boldsymbol{t}^{(k)}, \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\epsilon}_l^{(k)} - \boldsymbol{\mu}) \rangle) \\ & + \sin(\langle \boldsymbol{t}^{(k)}, \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\epsilon}_l^{(k)} - \boldsymbol{\mu}) \rangle) - \phi(\boldsymbol{t}^{(k)}) \Big] . \end{aligned}$$

Letting $\mu(\boldsymbol{t}; \boldsymbol{\gamma}) = E_{(\mathbf{0}, \boldsymbol{I})} g(\boldsymbol{\epsilon}_l, \boldsymbol{t}; \boldsymbol{\gamma})$, it is seen that $T_{n,b,A}$ is a V-statistic which falls into case I situation. In de Wet and Randles (1987), they refer to case I when all first order partial derivatives of $\mu(\boldsymbol{t}; \boldsymbol{\gamma})$ evaluated at the true parameter $\boldsymbol{\gamma} = \boldsymbol{\lambda}$ vanish. Otherwise, they refer to case II. This is case I here since only $A$'s such that $|A| > 1$ are considered. Thus, the asymptotic distribution of $T_{n,b,A}$ is the same whether one uses $\hat{\boldsymbol{\lambda}}_n$ or $\boldsymbol{\lambda}$ in (2.16). It is not clear, however, how this argument would apply to the joint distribution of $T_{n,b,A}$ and $T_{n,b,B}$. The proof of Theorem 2.4 in the Appendix does not use de Wet and Randles (1987).

For subsets $A$, $|A| = 1$, the statistic $T_{n,b,A}$ reduces to the statistic used by Baringhaus and Henze (1988) and Henze and Zirkler (1990) to test normality of a given marginal. They showed that the asymptotic distribution is affected by the estimation of the unknown parameters by establishing case II of de Wet and Randles (1987). Henze and Wagner (1997) treated the same problem with an approach based on stochastic processes.

### 2.4. Consistency

Consider the alternatives whereby $\boldsymbol{\epsilon}^{(1)}, \ldots, \boldsymbol{\epsilon}^{(p)}$ are distributed as $N_q(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, but are not independent. Then, $S_n \to \infty$ and $M_n \to \infty$. Thus, the test statistics $S_n$ and $M_n$ in (2.14)-(2.15) are consistent against any such alternatives.

The argument to establish consistency is rather trivial as in Baringhaus and Henze (1988). Recall that $C_p(\cdot)$ is the joint characteristic function of $\boldsymbol{\epsilon}^{(1)}, \ldots, \boldsymbol{\epsilon}^{(p)}$. This argument consists of the following almost sure convergence:

$$
\begin{aligned}
T_{n,b,A} &= \int |\sum_{B\subset A}(-1)^{|A\setminus B|}\exp(-\mathrm{i}\sum_{k\in B}\langle \boldsymbol{t}^{(k)}, \boldsymbol{S}^{-\frac{1}{2}}\overline{\boldsymbol{\epsilon}}\rangle) \\
&\quad \times\frac{1}{n}\sum_{j=1}^{n}\exp(\mathrm{i}\sum_{k\in B}\langle \boldsymbol{t}^{(k)}, \boldsymbol{S}^{-\frac{1}{2}}\boldsymbol{\epsilon}_j^{(k)}\rangle)\prod_{i\in A\setminus B}\phi(\boldsymbol{t}^{(i)})|^2\varphi_b(\boldsymbol{t})\mathrm{d}\boldsymbol{t} \\
&\stackrel{as}{\to} \int |\sum_{B\subset A}(-1)^{|A\setminus B|}\exp(-\mathrm{i}\sum_{k\in B}\langle \boldsymbol{t}^{(k)}, \boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\mu}\rangle) \\
&\quad \times C_p((\boldsymbol{I}_p\otimes\boldsymbol{\Sigma}^{-\frac{1}{2}})\boldsymbol{t}^B)\prod_{i\in A\setminus B}\phi(\boldsymbol{t}^{(i)})|^2\varphi_b(\boldsymbol{t})\mathrm{d}\boldsymbol{t}
\end{aligned}
$$

which equals 0 for all $A$, $|A|>1$, if and only if $\boldsymbol{\epsilon}^{(1)},\ldots,\boldsymbol{\epsilon}^{(p)}$ are independent $N_q(\boldsymbol{\mu},\boldsymbol{\Sigma})$. Therefore, if $\boldsymbol{\epsilon}^{(1)},\ldots,\boldsymbol{\epsilon}^{(p)}$ are dependent $N_q(\boldsymbol{\mu},\boldsymbol{\Sigma})$ then, there exists an $A$ such that $nT_{n,b,A}\to\infty$ which suffices to have $S_n\to\infty$ and $M_n\to\infty$.

## 3. Testing independence: the serial situation

### 3.1. The case of known parameters

Let $\boldsymbol{u}_1,\boldsymbol{u}_2,\ldots$ be a stationary sequence of random vectors $\boldsymbol{u}_i$ distributed as $N_q(\boldsymbol{0},\boldsymbol{I})$. It is desired to verify whether the $\boldsymbol{u}_i$'s are independent. Introduce the partitioned random vectors $\boldsymbol{\epsilon}_i=(\boldsymbol{u}_i,\ldots,\boldsymbol{u}_{i+p-1})\in\mathbb{R}^{pq}$, $i=1,\ldots,n-p+1$. Also, let $R_{n,A}(\boldsymbol{t})$ be as in (2.1) with the slight modification $\phi_{n,p}(\boldsymbol{t})=\frac{1}{n}\sum_{j=1}^{n-p+1}\exp(\mathrm{i}\langle\boldsymbol{t},\boldsymbol{\epsilon}_j\rangle)$. The main result related to the asymptotic distribution is that the $m$-dependence introduced by the overlapping of the $\boldsymbol{u}_i$'s does not affect the asymptotic distribution. It is the same as in the non-serial case.

**Theorem 3.1.** *If the $\boldsymbol{u}_i$'s are independent, the collection of processes $\{R_{n,A}:|A|>1\}$ converge in $C(\mathbb{R}^{pq},\mathbb{C})$ to independent zero mean complex Gaussian processes $\{R_A:|A|>1\}$ having covariance and pseudo-covariance functions respectively given by $C_A(\boldsymbol{s},\boldsymbol{t})$ and $\overline{C}_A(\boldsymbol{s},\boldsymbol{t})$ in (2.3) and (2.4).*

As in (2.5), the multinomial formula (Ghoudi et al. (2001)) yields

$$
R_{n,A}(\boldsymbol{t})=\frac{1}{\sqrt{n}}\sum_{j=1}^{n-p+1}\prod_{k\in A}[\exp(\mathrm{i}\langle\boldsymbol{t}^{(k)},\boldsymbol{\epsilon}_j^{(k)}\rangle)-\phi(\boldsymbol{t}^{(k)})]. \tag{3.1}
$$

which is useful in the proof of Theorem 3.1.

### 3.2. The case of unknown parameters

The context is the same as in the preceeding section but here the $\boldsymbol{u}_i$'s all have the same $N_q(\boldsymbol{\mu},\boldsymbol{\Sigma})$ normal distribution, where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, positive definite, are assumed unknown. Again, we want to test whether the $\boldsymbol{u}_i$'s are independent. To this aim, define the random vectors $\boldsymbol{\epsilon}_i=(\boldsymbol{u}_i,\ldots,\boldsymbol{u}_{i+p-1})\in\mathbb{R}^{pq}$ and $\boldsymbol{e}_i=(\hat{\boldsymbol{u}}_i,\ldots,\hat{\boldsymbol{u}}_{i+p-1})\in\mathbb{R}^{pq}$, $i=1,\ldots,n-p+1$. Also, define the standardized residuals $\hat{\boldsymbol{u}}_i=\boldsymbol{S}^{-\frac{1}{2}}(\boldsymbol{u}_i-\overline{\boldsymbol{u}})$ with the sample covariance matrix $\boldsymbol{S} = \frac{1}{n}\sum_{j=1}^{n}(\boldsymbol{u}_j - \overline{\boldsymbol{u}})(\boldsymbol{u}_j - \overline{\boldsymbol{u}})^{\mathsf{T}}$ and the sample mean $\overline{\boldsymbol{u}} = \frac{1}{n}\sum_{j=1}^{n}\boldsymbol{u}_j$. Now, let

$$\hat{R}_{n,A}(\boldsymbol{t}) = \sqrt{n}\sum_{B\subset A}(-1)^{|A\setminus B|}\hat{\phi}_{n,p}(\boldsymbol{t}^{B})\prod_{i\in A\setminus B}\phi(\boldsymbol{t}^{(i)}), \tag{3.2}$$

where $\hat{\phi}_{n,p}(\boldsymbol{t}) = \frac{1}{n}\sum_{j=1}^{n-p+1}\exp(\mathrm{i}\langle\boldsymbol{t},\boldsymbol{e}_j\rangle)$. The asymptotic behaviour of these processes is stated next. The main conclusion is that the estimation of the unknown parameters does not affect the asymptotic distribution.

**Theorem 3.2.** *If the $\mathbf{u}_i$'s are independent, the processes $\{\hat{R}_{n,A} : |A| > 1\}$ converge in $C(\mathbb{R}^{pq},\mathbb{C})$ to independent zero mean complex Gaussian processes $\{R_A : |A| > 1\}$ having covariance and pseudo-covariance functions respectively given by $C_A(\boldsymbol{s},\boldsymbol{t})$ and $\overline{C}_A(\boldsymbol{s},\boldsymbol{t})$ in (2.3) and (2.4).*

Note that the multinomial formula yields

$$\hat{R}_{n,A}(\boldsymbol{t}) = \frac{1}{\sqrt{n}}\sum_{j=1}^{n-p+1}\prod_{k\in A}[\exp(\mathrm{i}\langle\boldsymbol{t}^{(k)},\boldsymbol{e}_j^{(k)}\rangle) - \phi(\boldsymbol{t}^{(k)})] \tag{3.3}$$

and so the Cramér-von Mises test statistic

$$T_{n,b,A} = \frac{1}{n}\int|\hat{R}_{n,A}(\boldsymbol{t})|^2\varphi_b(\boldsymbol{t})\mathrm{d}\boldsymbol{t}$$

can be computed as

$$\frac{1}{n^2}\sum_{l=1}^{n-p+1}\sum_{l'=1}^{n-p+1}\prod_{k\in A}\Bigg\{\exp\left[-\frac{b^2}{2}\|\boldsymbol{e}_l^{(k)} - \boldsymbol{e}_{l'}^{(k)}\|^2\right] \tag{3.4}$$

$$-(b^2+1)^{-\frac{q}{2}}\exp\left[-\frac{1}{2}\frac{b^2}{b^2+1}\|\boldsymbol{e}_l^{(k)}\|^2\right] \tag{3.5}$$

$$-(b^2+1)^{-\frac{q}{2}}\exp\left[-\frac{1}{2}\frac{b^2}{b^2+1}\|\boldsymbol{e}_{l'}^{(k)}\|^2\right] + (2b^2+1)^{-\frac{q}{2}}\Bigg\}. \tag{3.6}$$

This representation shows that $T_{n,b,A}$ is affine invariant. Here again we can use theorem 2.3 to obtain

**Theorem 3.3.**

$$\int\left(|\hat{R}_{n,A}(\boldsymbol{t})|^2, |\hat{R}_{n,B}(\boldsymbol{t})|^2\right)\varphi_b(\boldsymbol{t})d\boldsymbol{t} \xrightarrow{\mathcal{D}} \int\left(|R_A(\boldsymbol{t})|^2, |R_B(\boldsymbol{t})|^2\right)\varphi_b(\boldsymbol{t})d\boldsymbol{t}, \tag{3.7}$$

*where integrals are computed componentwise.*

In the serial situation, a subset $A$ and its translate $A+k$ essentially lead to the same statistic $T_{n,b,A}$. Hence, when considering these statistics, only $A$'s such that $1\in A$ can be considered. The same statistics (2.14) or (2.15) can be used to perform the statistical test:

$$S_n = n\sum_{|A|>1,1\in A}T_{n,b,A}, \qquad M_n = n\max_{|A|>1,1\in A}T_{n,b,A}.$$

# 4. PROPERTIES OF THE LIMITING PROCESSES

This section shows how to compute the critical values of the Cramér-von Mises variable $T_{b,A} \equiv \int |R_A(\boldsymbol{t})|^2 \varphi_b(\boldsymbol{t})\mathrm{d}\boldsymbol{t}$. This can be achieved either by computing its cumulants and then applying the Cornish-Fisher asymptotic expansion (see Lee and Lin (1992); Lee and Lee (1992)) or by inversion of the characteristic function (see Imhof (1961) or an improved version of this algorithm introduced by Davies (1973, 1980) or Deheuvels and Martynov (1996)) after evaluation of the eigenvalues of $C_A$.

The Cramér-von Mises test statistic in (2.9)can also be written in terms of a real process:

$$T_{n,b,A} = \frac{1}{n}\int W_{n,A}^2(\boldsymbol{t})\varphi_b(\boldsymbol{t})\mathrm{d}\boldsymbol{t}$$

where $W_{n,A}(\boldsymbol{t}) = \frac{1}{\sqrt{n}}\sum_{j=1}^n \prod_{k\in A}\left[\cos(\langle \boldsymbol{t}^{(k)}, \boldsymbol{e}_j^{(k)}\rangle) + \sin(\langle \boldsymbol{t}^{(k)}, \boldsymbol{e}_j^{(k)}\rangle) - \phi(\boldsymbol{t}^{(k)})\right]$ is a real process which converges to a real Gaussian process with the same covariance function $C_A(\boldsymbol{s}, \boldsymbol{t})$ as in (2.3). Thus the usual Karhunen-Loève expansion holds.

Let $k = |A|$. It is well known that

$$\int_{\mathbb{R}^{pq}} |R_A(\boldsymbol{t})|^2\varphi_b(\boldsymbol{t})\mathrm{d}\boldsymbol{t} \sim T_{b,A} = \sum_{(i_1,\ldots,i_k)\in\mathbb{N}^{*k}} \lambda_{(i_1,\ldots,i_k)}|Z_{(i_1,\ldots,i_k)}|^2 \tag{4.1}$$

where $Z_{(i_1,\ldots,i_k)}$ are standard i.i.d. $CN(0,1)$ complex random variables. Also, it is easy to show that $\lambda_{(i_1,\ldots,i_k)} = \prod_{l=1}^k \lambda_{i_l}$ where the $\lambda_j$'s are the eigenvalues of the integral operator $O$ defined by

$$O(f)(\boldsymbol{s}) = \int_{\mathbb{R}^q} f(\boldsymbol{t})K(\boldsymbol{s},\boldsymbol{t})\varphi_b(\boldsymbol{t})\mathrm{d}\boldsymbol{t} \tag{4.2}$$

with

$$K(\boldsymbol{s},\boldsymbol{t}) = \exp(-\frac{1}{2}\|\boldsymbol{s}-\boldsymbol{t}\|^2) - \exp(-\frac{1}{2}(\|\boldsymbol{s}\|^2 + \|\boldsymbol{t}\|^2)). \tag{4.3}$$

That is to say the problem is to solve, in $\lambda$ (and $f$), the linear second order homogeneous Fredholm integral equation

$$\lambda f(\boldsymbol{s}) = \int_{\mathbb{R}^q} f(\boldsymbol{t})K(\boldsymbol{s},\boldsymbol{t})\varphi_b(\boldsymbol{t})\mathrm{d}\boldsymbol{t}. \tag{4.4}$$

See Conway (1985) for an introduction to integral operators.

It does not seem possible to solve (4.4) explicitely, but one can compute its eigenvalues using a relation as

$$\int_{\mathbb{R}^q} f(\boldsymbol{\nu})e^{-\|\boldsymbol{\nu}\|^2}\mathrm{d}\boldsymbol{\nu} = \sum_{j=1}^N B_j f(\boldsymbol{\nu}_j) \tag{4.5}$$

where the parameters $B_j$ and $\boldsymbol{\nu}_j = (\nu_{j,1},\ldots,\nu_{j,q})$, $j = 1,\ldots,N$ could be respectively the coefficients and the points of a cubature formula (CF), or also could be obtained by a Monte-Carlo experiment, in which case $B_j = \frac{1}{N}$ and $\boldsymbol{\nu}_j \sim N(\boldsymbol{0}, \boldsymbol{I})$, $j = 1,\ldots,N$. A good rule of the thumb is to use a cubature formula when $b$ is small, for example less than one, otherwise use Monte Carlo method.

We used the following formulas: the $n$th degree Gauss quadrature formula when $q = 1$, the fifteenth degree CF $E_2^{r^2} : 15 - 1$ (see (Stroud, 1971, p. 326)) when $q = 2$

and the seventh degree CF $E_q^{r^2} : 7-2$ appearing in (Stroud, 1971, p. 319) for $q \geq 3$. This last formula contains an error, see Stroud (1967) for details.

Moreover one can notice that all the cumulants of $T_{b,A}$ can be computed explicitly. In fact, the $m$-th cumulant $\kappa_{m,A}(b)$ of $T_{b,A}$ in (4.1) is given by

$$\kappa_{m,A}(b) = (m-1)! \left[ \int_{\mathbb{R}^q} K^{(m)}(\boldsymbol{x}, \boldsymbol{x}) \varphi_b(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \right]^{|A|} \tag{4.6}$$

where

$$\varphi_b(\boldsymbol{x}) = (2\pi)^{-q/2} b^{-q} \exp(-\frac{1}{2} \|\boldsymbol{x}\|^2 / b^2) \tag{4.7}$$

and where $K^{(1)}(\boldsymbol{x}, \boldsymbol{y}) = K(\boldsymbol{x}, \boldsymbol{y})$ and

$$K^{(m)}(\boldsymbol{x}, \boldsymbol{y}) = \int_{\mathbb{R}^q} K^{(m-1)}(\boldsymbol{x}, \boldsymbol{z}) K(\boldsymbol{z}, \boldsymbol{y}) \varphi_b(\boldsymbol{z}) \mathrm{d}\boldsymbol{z},\ m \geq 2.$$

Define

$$\begin{aligned} I^{(1)}_{\boldsymbol{x},\boldsymbol{y}}(\alpha, \beta, \gamma) &= \exp(\alpha \|\boldsymbol{x}\|^2 + \beta \|\boldsymbol{y}\|^2 + \gamma \langle \boldsymbol{x}, \boldsymbol{y} \rangle) \\ I^{(m)}_{\boldsymbol{x},\boldsymbol{y}}(\alpha, \beta, \gamma) &= \int_{\mathbb{R}^q} I^{(m-1)}_{\boldsymbol{x},\boldsymbol{z}}(\alpha, \beta, \gamma) K(\boldsymbol{z}, \boldsymbol{y}) \varphi_b(\boldsymbol{z}) \mathrm{d}\boldsymbol{z},\ m \geq 2. \end{aligned}$$

One can show

$$K^{(m)}(\boldsymbol{x}, \boldsymbol{x}) = I^{(m)}_{\boldsymbol{x},\boldsymbol{x}} \left( -\frac{1}{2}, -\frac{1}{2}, 1 \right) - I^{(m)}_{\boldsymbol{x},\boldsymbol{x}} \left( -\frac{1}{2}, -\frac{1}{2}, 0 \right) \tag{4.8}$$

and the recurrence relation

$$I^{(m)}_{\boldsymbol{x},\boldsymbol{x}}(\alpha, \beta, \gamma) = \chi_\beta \left[ I^{(m-1)}_{\boldsymbol{x},\boldsymbol{x}}(\alpha - \frac{\gamma^2}{4c}, -\frac{1}{4c} - \frac{1}{2}, -\frac{2\gamma}{4c}) - I^{(m-1)}_{\boldsymbol{x},\boldsymbol{x}}(\alpha - \frac{\gamma^2}{4c}, -\frac{1}{2}, 0) \right],$$

where $\chi_\beta = (2\pi)^{-q/2} b^{-q} \int_{\mathbb{R}^q} e^{c\|\boldsymbol{z}\|^2} \mathrm{d}\boldsymbol{z} = (1 + b^2 - 2b^2\beta)^{-q/2}$ and $-\frac{1}{4c} = \frac{b^2}{2+2b^2-4b^2\beta}$. Thus, one can express $K^{(m)}(\boldsymbol{x}, \boldsymbol{x})$ in terms of $I^{(1)}_{\boldsymbol{x},\boldsymbol{x}}$ and then use the relation

$$\int_{\mathbb{R}^q} I^{(1)}_{\boldsymbol{x},\boldsymbol{x}}(\alpha, \beta, \gamma) \varphi_b(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} = [1 - 2b^2(\alpha + \beta + \gamma)]^{-q/2} \tag{4.9}$$

to obtain all the cumulants recursively. Note that this permits to double-check into the preceding computation of the eigenvalues through the following relation

$$\kappa_{m,A}(b) = (m-1)! \left[ \sum_{j=1}^{\infty} \lambda_j^m \right]^{|A|}. \tag{4.10}$$

Note also that one only needs to compute the cumulants $\kappa_m(b)$'s for the case $|A| = 1$ since

$$\kappa_{m,A}(b) = [\kappa_m(b)]^{|A|} \left[ (m-1)! \right]^{1-|A|}. \tag{4.11}$$

The CF's used here are not the only one available to obtain estimates $\hat{\lambda}_j$ of the $\lambda_j$'s. A good choice is one that minimises

$$|\kappa_m(b) - (m-1)! \sum_{j=1}^{\infty} \hat{\lambda}_j^m|. \tag{4.12}$$

See Cools (1999) and Cools and Rabinowitz (1993) for a comprehensive list of such formulas.
Table 4.1 provides an approximation of the cut-off values obtained from the Cornish Fisher asymptotic expansion based on the first six cumulants, for $b = 0.1$.

TABLE 4.1. Critical Values of the Distribution of $T_{b,A}$ for $b = 0.1$.

| $1-\alpha$ | q=2 | | | q=3 | | |
|---|---|---|---|---|---|---|
| | k=2 | k=3 | k=4 | k=2 | k=3 | k=4 |
| 0.900 | 0.000733 | 1.230e-05 | 2.122e-07 | 0.00137 | 3.347e-05 | 8.707e-07 |
| 0.905 | 0.000745 | 1.244e-05 | 2.140e-07 | 0.00138 | 3.370e-05 | 8.741e-07 |
| 0.910 | 0.000758 | 1.260e-05 | 2.159e-07 | 0.00140 | 3.393e-05 | 8.777e-07 |
| 0.915 | 0.000771 | 1.276e-05 | 2.179e-07 | 0.00142 | 3.418e-05 | 8.815e-07 |
| 0.920 | 0.000785 | 1.293e-05 | 2.200e-07 | 0.00143 | 3.444e-05 | 8.854e-07 |
| 0.925 | 0.000800 | 1.311e-05 | 2.222e-07 | 0.00145 | 3.472e-05 | 8.896e-07 |
| 0.930 | 0.000815 | 1.330e-05 | 2.246e-07 | 0.00147 | 3.501e-05 | 8.939e-07 |
| 0.935 | 0.000832 | 1.350e-05 | 2.271e-07 | 0.00149 | 3.531e-05 | 8.986e-07 |
| 0.940 | 0.000850 | 1.372e-05 | 2.298e-07 | 0.00152 | 3.564e-05 | 9.035e-07 |
| 0.945 | 0.000870 | 1.395e-05 | 2.326e-07 | 0.00154 | 3.600e-05 | 9.088e-07 |
| 0.950 | 0.000891 | 1.421e-05 | 2.357e-07 | 0.00157 | 3.638e-05 | 9.145e-07 |
| 0.955 | 0.000915 | 1.449e-05 | 2.392e-07 | 0.00160 | 3.680e-05 | 9.207e-07 |
| 0.960 | 0.000941 | 1.480e-05 | 2.429e-07 | 0.00163 | 3.726e-05 | 9.275e-07 |
| 0.965 | 0.000971 | 1.514e-05 | 2.471e-07 | 0.00167 | 3.777e-05 | 9.351e-07 |
| 0.970 | 0.001005 | 1.554e-05 | 2.519e-07 | 0.00171 | 3.836e-05 | 9.437e-07 |
| 0.975 | 0.001045 | 1.601e-05 | 2.575e-07 | 0.00176 | 3.904e-05 | 9.537e-07 |
| 0.980 | 0.001093 | 1.657e-05 | 2.643e-07 | 0.00182 | 3.985e-05 | 9.656e-07 |
| 0.985 | 0.001156 | 1.729e-05 | 2.728e-07 | 0.00190 | 4.088e-05 | 9.805e-07 |
| 0.990 | 0.001243 | 1.828e-05 | 2.845e-07 | 0.00200 | 4.229e-05 | 1.000e-06 |
| 0.995 | 0.001390 | 1.994e-05 | 3.039e-07 | 0.00217 | 4.460e-05 | 1.033e-06 |

TABLE 4.2. Distribution of $T^*_{b,A}$ for $b = 0.1$.

| x | q=2 | | | q=3 | | |
|---|---|---|---|---|---|---|
| | $P[T_{b,2} \leq x]$ | $P[T_{b,3} \leq x]$ | $P[T_{b,4} \leq x]$ | $P[T_{b,2} \leq x]$ | $P[T_{b,3} \leq x]$ | $P[T_{b,4} \leq x]$ |
| 0.0 | 0.593 | 0.566 | 0.546 | 0.562 | 0.536 | 0.520 |
| 0.2 | 0.665 | 0.640 | 0.622 | 0.637 | 0.612 | 0.598 |
| 0.4 | 0.725 | 0.705 | 0.691 | 0.702 | 0.683 | 0.671 |
| 0.6 | 0.777 | 0.761 | 0.751 | 0.759 | 0.745 | 0.737 |
| 0.8 | 0.819 | 0.809 | 0.802 | 0.807 | 0.798 | 0.794 |
| 1.0 | 0.854 | 0.848 | 0.845 | 0.848 | 0.843 | 0.842 |
| 1.2 | 0.883 | 0.881 | 0.880 | 0.880 | 0.880 | 0.881 |
| 1.4 | 0.906 | 0.907 | 0.908 | 0.907 | 0.910 | 0.913 |
| 1.6 | 0.925 | 0.928 | 0.931 | 0.928 | 0.933 | 0.937 |
| 1.8 | 0.941 | 0.944 | 0.948 | 0.945 | 0.951 | 0.955 |
| 2.0 | 0.953 | 0.957 | 0.961 | 0.958 | 0.964 | 0.969 |
| 2.2 | 0.963 | 0.967 | 0.972 | 0.968 | 0.974 | 0.979 |
| 2.4 | 0.970 | 0.975 | 0.979 | 0.976 | 0.982 | 0.986 |
| 2.6 | 0.977 | 0.981 | 0.985 | 0.982 | 0.987 | 0.990 |
| 2.8 | 0.982 | 0.986 | 0.989 | 0.986 | 0.991 | 0.994 |
| 3.0 | 0.985 | 0.989 | 0.992 | 0.990 | 0.994 | 0.996 |
| 3.2 | 0.988 | 0.992 | 0.994 | 0.992 | 0.996 | 0.997 |
| 3.4 | 0.991 | 0.994 | 0.996 | 0.994 | 0.997 | 0.998 |
| 3.6 | 0.993 | 0.995 | 0.997 | 0.996 | 0.998 | 0.999 |
| 3.8 | 0.994 | 0.996 | 0.998 | 0.997 | 0.998 | 0.999 |
| 4.0 | 0.995 | 0.997 | 0.998 | 0.997 | 0.999 | 0.999 |

As in Ghoudi et al. (2001), define $T^*_{b,A}$ to be the standardised version of $T_{b,A}$. Then Table 4.2 provides the distribution function of this statistic for some values of $|A|$ and $q$, with $b = 0.1$ as approximated by Davies technique. Besides, a C++ program is available from the authors which permits to compute any cut-off value given the nominal level and vice-versa.

In Table 4.3, one can find the empirical percentage points of $nT_{n,b,A}$ $(n = 20, 50, 100; b = 0.1, 0.5, 1.0, 3.0; \alpha = 0.1, 0.05)$ based on $N = 10000$ Monte Carlo replications, in the non serial case.

TABLE 4.3. Empirical Percentage Points of $nT_{n,b,A}$ based on $N = 10000$ Monte Carlo replications: non-serial case.

| $1-\alpha$ | $b$ | $n$ | q=2 | | | q=3 | | |
|---|---|---|---|---|---|---|---|---|
| | | | $k=2$ | $k=3$ | $k=4$ | $k=2$ | $k=3$ | $k=4$ |
| 0.9 | 0.1 | 20 | 0.000718 | 1.249e-05 | 2.371e-07 | 0.00134 | 3.460e-05 | 1.033e-06 |
| | | 50 | 0.000718 | 1.253e-05 | 2.358e-07 | 0.00136 | 3.420e-05 | 9.748e-07 |
| | | 100 | 0.000733 | 1.229e-05 | 2.242e-07 | 0.00137 | 3.389e-05 | 9.430e-07 |
| | 0.5 | 20 | 0.170 | 0.049 | 0.016 | 0.269 | 0.110 | 0.050 |
| | | 50 | 0.167 | 0.0481 | 0.0156 | 0.266 | 0.107 | 0.0484 |
| | | 100 | 0.168 | 0.0471 | 0.0149 | 0.266 | 0.106 | 0.0472 |
| | 1.0 | 20 | 0.561 | 0.333 | 0.221 | 0.727 | 0.555 | 0.450 |
| | | 50 | 0.558 | 0.327 | 0.213 | 0.724 | 0.547 | 0.440 |
| | | 100 | 0.559 | 0.326 | 0.209 | 0.723 | 0.544 | 0.436 |
| | 3.0 | 20 | 0.938 | 0.864 | 0.820 | 0.981 | 0.969 | 0.958 |
| | | 50 | 0.940 | 0.860 | 0.814 | 0.983 | 0.967 | 0.955 |
| | | 100 | 0.938 | 0.858 | 0.811 | 0.983 | 0.966 | 0.954 |
| 0.95 | 0.1 | 20 | 0.000854 | 1.485e-05 | 3.036e-07 | 0.00151 | 3.849e-05 | 1.212e-06 |
| | | 50 | 0.000883 | 1.465e-05 | 2.840e-07 | 0.00156 | 3.751e-05 | 1.103e-06 |
| | | 100 | 0.000874 | 1.429e-05 | 2.653e-07 | 0.00155 | 3.724e-05 | 1.027e-06 |
| | 0.5 | 20 | 0.191 | 0.0539 | 0.0184 | 0.289 | 0.114 | 0.053 |
| | | 50 | 0.192 | 0.0520 | 0.0168 | 0.289 | 0.111 | 0.0505 |
| | | 100 | 0.191 | 0.0513 | 0.0158 | 0.289 | 0.110 | 0.0486 |
| | 1.0 | 20 | 0.606 | 0.345 | 0.228 | 0.749 | 0.562 | 0.456 |
| | | 50 | 0.603 | 0.338 | 0.218 | 0.748 | 0.553 | 0.445 |
| | | 100 | 0.593 | 0.336 | 0.213 | 0.745 | 0.549 | 0.439 |
| | 3.0 | 20 | 0.955 | 0.868 | 0.823 | 0.984 | 0.970 | 0.959 |
| | | 50 | 0.954 | 0.863 | 0.816 | 0.987 | 0.967 | 0.956 |
| | | 100 | 0.953 | 0.861 | 0.813 | 0.986 | 0.966 | 0.955 |

In Table 4.4, one can find the empirical percentage points of $nT_{n,b,A}$ $(n = 20, 50, 100; b = 0.1, 0.5, 1.0, 3.0; \alpha = 0.1, 0.05)$ based on $N = 10000$ Monte Carlo replications, in the serial case, for $p = 4$.

TABLE 4.4. Empirical Percentage Points of $nT_{n,b,A}$ based on $N = 10000$ Monte Carlo replications for $p = 4$: serial case.

| $1-\alpha$ | $b$ | $n$ | q=2 | | | q=3 | | |
|---|---|---|---|---|---|---|---|---|
| | | | $k=2$ | $k=3$ | $k=4$ | $k=2$ | $k=3$ | $k=4$ |
| 0.9 | 0.1 | 20 | 0.000712 | 1.174e-05 | 2.139e-07 | 0.00132 | 3.384e-05 | 9.931e-07 |
| | | 50 | 0.000721 | 1.205e-05 | 2.288e-07 | 0.00135 | 3.357e-05 | 9.867e-07 |
| | | 100 | 0.000716 | 1.217e-05 | 2.261e-07 | 0.00138 | 3.335e-05 | 9.530e-07 |
| | 0.5 | 20 | 0.173 | 0.0504 | 0.0171 | 0.276 | 0.1160 | 0.0547 |
| | | 50 | 0.168 | 0.0484 | 0.0162 | 0.269 | 0.109 | 0.050 |
| | | 100 | 0.170 | 0.0476 | 0.0155 | 0.268 | 0.107 | 0.048 |
| | 1.0 | 20 | 0.567 | 0.353 | 0.241 | 0.737 | 0.586 | 0.486 |
| | | 50 | 0.561 | 0.334 | 0.223 | 0.728 | 0.559 | 0.458 |
| | | 100 | 0.561 | 0.328 | 0.216 | 0.727 | 0.55 | 0.447 |
| | 3.0 | 20 | 0.937 | 0.880 | 0.842 | 0.984 | 0.975 | 0.966 |
| | | 50 | 0.939 | 0.867 | 0.825 | 0.983 | 0.970 | 0.960 |
| | | 100 | 0.939 | 0.862 | 0.818 | 0.983 | 0.968 | 0.957 |
| 0.95 | 0.1 | 20 | 0.000841 | 1.414e-05 | 2.734e-07 | 0.00148 | 3.804e-05 | 1.176e-06 |
| | | 50 | 0.000897 | 1.412e-05 | 2.780e-07 | 0.00154 | 3.772e-05 | 1.128e-06 |
| | | 100 | 0.000861 | 1.435e-05 | 2.737e-07 | 0.00154 | 3.688e-05 | 1.071e-06 |
| | 0.5 | 20 | 0.196 | 0.0556 | 0.0193 | 0.296 | 0.122 | 0.0587 |
| | | 50 | 0.192 | 0.0528 | 0.0179 | 0.292 | 0.114 | 0.0539 |
| | | 100 | 0.189 | 0.0517 | 0.0168 | 0.290 | 0.111 | 0.0509 |
| | 1.0 | 20 | 0.615 | 0.367 | 0.255 | 0.759 | 0.598 | 0.498 |
| | | 50 | 0.604 | 0.346 | 0.231 | 0.753 | 0.567 | 0.466 |
| | | 100 | 0.602 | 0.340 | 0.222 | 0.751 | 0.557 | 0.452 |
| | 3.0 | 20 | 0.955 | 0.886 | 0.849 | 0.985 | 0.976 | 0.968 |
| | | 50 | 0.953 | 0.871 | 0.829 | 0.987 | 0.971 | 0.961 |
| | | 100 | 0.952 | 0.865 | 0.822 | 0.987 | 0.969 | 0.958 |

## 5. ONE-WAY MANOVA MODEL WITH RANDOM EFFECTS

The one way linear model with random effects

$$\boldsymbol{\epsilon}_i^{(j)} = \boldsymbol{\mu} + \boldsymbol{\alpha}_i + \boldsymbol{\delta}_i^{(j)},\ i = 1, \ldots, n; j = 1, \ldots, p,$$

where $\boldsymbol{\alpha}_i \sim N_q(\mathbf{0}, \boldsymbol{\psi})$ and $\boldsymbol{\delta}_i^{(j)} \sim N_q(\mathbf{0}, \boldsymbol{\Sigma})$ are all mutually independant, provides a joint normal model for the non-serial case. This means that in this variance component model

$$\boldsymbol{\epsilon}_i = (\boldsymbol{\epsilon}_i^{(1)}, \ldots, \boldsymbol{\epsilon}_i^{(p)}) \sim N_{pq}(\mathbf{1}_p \otimes \boldsymbol{\mu}, (\boldsymbol{I}_p \otimes \boldsymbol{\Sigma}) + (\mathbf{1}_p \mathbf{1}_p^{\mathsf{T}}) \otimes \boldsymbol{\psi}), i = 1, \ldots, n,$$

are i.i.d. The test of independance amounts to the parametric test of the hypothesis $H_0 : \boldsymbol{\psi} = \mathbf{0}$. The MANOVA decomposition

$$\begin{aligned}
\sum_{i=1}^{n}\sum_{j=1}^{p}(\boldsymbol{\epsilon}_i^{(j)} - \overline{\boldsymbol{\epsilon}}_{\bullet}^{(\bullet)})(\boldsymbol{\epsilon}_i^{(j)} - \overline{\boldsymbol{\epsilon}}_{\bullet}^{(\bullet)})^{\mathsf{T}} &= \sum_{i=1}^{n}\sum_{j=1}^{p}(\boldsymbol{\epsilon}_i^{(j)} - \overline{\boldsymbol{\epsilon}}_{i}^{(\bullet)})(\boldsymbol{\epsilon}_i^{(j)} - \overline{\boldsymbol{\epsilon}}_{i}^{(\bullet)})^{\mathsf{T}} && (5.1)\\
&+ p\sum_{i=1}^{n}(\overline{\boldsymbol{\epsilon}}_i^{(\bullet)} - \overline{\boldsymbol{\epsilon}}_{\bullet}^{(\bullet)})(\overline{\boldsymbol{\epsilon}}_i^{(\bullet)} - \overline{\boldsymbol{\epsilon}}_{\bullet}^{(\bullet)})^{\mathsf{T}} && (5.2)\\
&= E + H, && (5.3)
\end{aligned}$$

(see Rencher (2002)), leads to the usual MANOVA table. A dot means averaging over the corresponding index. The test of $H_0 : \boldsymbol{\psi} = \mathbf{0}$ is usually done with Wilks statistic

TABLE 5.1. MANOVA table

| Source | Sum of squares | Degrees of freedom | Expected sum of squares | Null distribution |
|---|---|---|---|---|
| Between | $H$ | $\nu_H = n-1$ | $(n-1)\boldsymbol{\Sigma} + p(n-1)\boldsymbol{\psi}$ | $W_q(n-1, \boldsymbol{\Sigma})$ |
| Within | $E$ | $\nu_E = n(p-1)$ | $n(p-1)\boldsymbol{\Sigma}$ | $W_q(n(p-1), \boldsymbol{\Sigma})$ |
| Total | $E+H$ | | | |

$$\Lambda = \frac{\det E}{\det(E+H)} = \prod_{k=1}^{q} \left[1 + \lambda_k(E^{-1}H)\right]^{-1},$$

where $\lambda_k(E^{-1}H)$ are the eigenvalues of $E^{-1}H$. The null distribution of $\Lambda$ is the $\Lambda_{q,\nu_H,\nu_E}$ distribution. Tables of exact critical points for $\Lambda$ are available but for $q = 2$ the relation

$$\frac{(\nu_E - 1)}{\nu_H} \frac{(1 - \Lambda^{\frac{1}{2}})}{\Lambda^{\frac{1}{2}}} \sim F_{2\nu_H, 2(\nu_E - 1)}$$

holds.

TABLE 5.2. Empirical power of $nT_{n,b,A}$ and Wilks test based on $N = 10000$ Monte Carlo replications for $p = 2$, $q = 2$, $\boldsymbol{\mu} = \mathbf{0}$, $\boldsymbol{\Sigma} = \gamma \boldsymbol{I}_2$ and $\boldsymbol{\Psi} = \theta \boldsymbol{I}_2$.

| $1-\alpha$ | $\frac{\theta}{\gamma}$ | $n$ | $nT_{n,b,A}$ | | | | | Wilks | $\frac{\lvert\hat{R}_{n,A}(\boldsymbol{t})\rvert^2}{0.5C_A(\boldsymbol{t},\boldsymbol{t})}$ |
|---|---|---|---|---|---|---|---|---|---|
| | | | $b = 0.01$ | $b = 0.1$ | $b = 0.5$ | $b = 1.0$ | $b = 3.0$ | | $\boldsymbol{t} = (0.6, 0.6, 0.6, 0.6)$ |
| 0.90 | 0.0 | 20 | 10.14 | 10.27 | 10.06 | 10.21 | 10.33 | 9.96 | 10.64 |
| | | 50 | 9.46 | 10.14 | 10.15 | 10.20 | 9.94 | 10.34 | 9.99 |
| | 0.2 | 20 | 14.27 | 14.39 | 13.78 | 12.64 | 11.12 | 39.15 | 22.23 |
| | | 50 | 28.61 | 29.87 | 27.50 | 19.00 | 10.50 | 63.63 | 29.44 |
| | 0.4 | 20 | 28.43 | 28.57 | 26.61 | 19.28 | 11.81 | 67.69 | 35.55 |
| | | 50 | 68.31 | 69.70 | 63.43 | 40.63 | 13.67 | 94.72 | 49.78 |
| | 1.0 | 20 | 77.09 | 77.36 | 72.66 | 50.48 | 16.49 | 97.86 | 51.98 |
| | | 50 | 99.80 | 99.81 | 99.43 | 92.68 | 27.66 | 99.99 | 60.26 |
| 0.95 | 0.0 | 20 | 4.67 | 4.91 | 5.52 | 4.88 | 5.16 | 5.22 | 5.39 |
| | | 50 | 4.84 | 4.58 | 4.98 | 4.82 | 5.04 | 5.22 | 5.16 |
| | 0.2 | 20 | 7.35 | 7.63 | 7.94 | 6.39 | 5.53 | 26.10 | 13.86 |
| | | 50 | 18.96 | 18.26 | 17.11 | 10.60 | 5.14 | 49.60 | 20.06 |
| | 0.4 | 20 | 16.77 | 17.20 | 16.93 | 10.80 | 6.03 | 53.44 | 25.04 |
| | | 50 | 56.83 | 55.72 | 50.23 | 28.08 | 7.37 | 89.24 | 38.59 |
| | 1.0 | 20 | 64.57 | 65.61 | 61.12 | 36.17 | 8.98 | 94.91 | 42.47 |
| | | 50 | 99.45 | 99.42 | 98.61 | 86.81 | 17.72 | 99.98 | 50.93 |

Wilks test has power superior to the test proposed in this paper. This is not surprising since Wilks test is specifically designed for the parametric hypothesis $\boldsymbol{\Psi} = \mathbf{0}$ in the linear model which holds in the simulation. However, our test is more generally applicable and will yield reasonable power for reasonably large sample sizes. Moreover, it is not difficult to construct another model where Wilks test would fail. For example, Wilks test would be unable to detect the dependance in the normal mixture model

$$\begin{pmatrix} \boldsymbol{\epsilon}^{(1)} \\ \boldsymbol{\epsilon}^{(2)} \end{pmatrix} \sim 0.5 N_4 \left( \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{I}_2 & \boldsymbol{\Psi} \\ \boldsymbol{\Psi} & \boldsymbol{I}_2 \end{pmatrix} \right) + 0.5 N_4 \left( \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{I}_2 & -\boldsymbol{\Psi} \\ -\boldsymbol{\Psi} & \boldsymbol{I}_2 \end{pmatrix} \right)$$

contrary to our test.

## 6. Proofs

Define the metric $\rho$ on $C := C(\mathbb{R}^{pq}, \mathbb{C})$ by

$$\rho(x, y) = \sum_{j=1}^{\infty} 2^{-j} \frac{\rho_j(x, y)}{1 + \rho_j(x, y)}, \tag{6.1}$$

where

$$\rho_j(x, y) = \sup_{\|\boldsymbol{t}\| \leq j} |x(\boldsymbol{t}) - y(\boldsymbol{t})|$$

is the usual sup norm. Endowed with this metric, $C(\mathbb{R}^p, \mathbb{C})$ is a separable Fréchet space, that is a linear complete and separable metric space. Moreover, convergence in this metric corresponds to the uniform convergence on all compact sets, $\lim_{n\to\infty} \rho(x_n, x) = 0$ if and only if $\forall j \geq 1$, $\lim_{n\to\infty} \rho_j(x_n, x) = 0$. For random elements, it also holds $\rho(X_n, Y_n) \xrightarrow{P} 0$ if and only if $\forall j \geq 1, \rho_j(X_n, Y_n) \xrightarrow{P} 0$. For all mappings $x \in C$, let $r_j(x)$ be the restriction of $x$ to the ball $\mathbb{R}^{pq}_j$ of radius $j$.

### 6.1. Proof of Theorem 2.1

By Propositions 14.6, 14.2 and Theorem 14.3 of Kallenberg (1997), it suffices to show that $r_j(R_{n,A}) \xrightarrow{fd} r_j(R_A)$ and $\{r_j(R_{n,A})\}$ is a tight family in order to show the convergence $R_{n,A} \xrightarrow{\mathcal{D}} R_A$. Convergence of the finite dimensional distributions to Gaussian limit is a direct consequence of the multivariate central limit theorem and the representation (2.5) of the process. The covariance function $C_A$ of (2.3) is also easy to obtain, as is the independence of the processes $R_A$. Then, one can write

$$R_{n,A}(\boldsymbol{t}) = \sum_{B \subset A} (-1)^{|A \setminus B|} \phi_{A,B}(\boldsymbol{t}) Y_{n,B}(\boldsymbol{t}), \tag{6.2}$$

where

$$Y_{n,B}(\boldsymbol{t}) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left\{ \exp(\mathrm{i}\langle \boldsymbol{t}^B, \boldsymbol{\epsilon}_j \rangle) - E\left[ \exp(\mathrm{i}\langle \boldsymbol{t}^B, \boldsymbol{\epsilon}_1 \rangle) \right] \right\} \tag{6.3}$$

and $\phi_{A,B}(\boldsymbol{t}) = \prod_{l \in A \setminus B} \phi(\boldsymbol{t}^{(l)})$. The process $Y_{n,B}$ on a compact is an empirical characteristic function process which was shown by Csörgő (1981b) to be tight if the sufficient condition in (Csörgő, 1985, p. 294) is satisfied. Hence, $r_j(Y_{n,B})$ is tight. Since there is only a finite number of $B$'s in (6.2), it follows that $\{r_j(R_{n,A})\}$ is also tight.

### 6.2. Proof of Theorem 2.2

By invariance assume $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma} = \boldsymbol{I}$. Let $\boldsymbol{e}_j^{(k)} = \boldsymbol{S}^{-\frac{1}{2}}(\boldsymbol{\epsilon}_j^{(k)} - \overline{\boldsymbol{\epsilon}}) = \boldsymbol{\epsilon}_j^{(k)} + \boldsymbol{\Delta}_j^{(k)}$, with $\boldsymbol{\Delta}_j^{(k)} = (\boldsymbol{S}^{-\frac{1}{2}} - \boldsymbol{I})\boldsymbol{\epsilon}_j^{(k)} - \boldsymbol{S}^{-\frac{1}{2}}\overline{\boldsymbol{\epsilon}}$. Two intermediate processes are now defined. Firstly,

$$\tilde{R}_{n,A}(\boldsymbol{t}) = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left\{\prod_{k\in A}(\exp(\mathrm{i}\langle \boldsymbol{t}^{(k)}, \boldsymbol{\epsilon}_j^{(k)}\rangle) - \phi(\boldsymbol{t}^{(k)}))\right. \tag{6.4}$$

$$\left. +\mathrm{i}\sum_{\alpha\in A} a_{\alpha,j} \prod_{k\in A;k\neq\alpha}\left[\exp(\mathrm{i}\langle \boldsymbol{t}^{(k)}, \boldsymbol{\epsilon}_j^{(k)}\rangle) - \phi(\boldsymbol{t}^{(k)})\right]\exp(\mathrm{i}\langle \boldsymbol{t}^{(\alpha)}, \boldsymbol{\epsilon}_j^{(\alpha)}\rangle)\right\} \tag{6.5}$$

$$\equiv \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\prod_{k\in A}[\exp(\mathrm{i}\langle \boldsymbol{t}^{(k)}, \boldsymbol{\epsilon}_j^{(k)}\rangle) - \phi(\boldsymbol{t}^{(k)})] + \mathrm{i}U_n(\boldsymbol{t}) \tag{6.6}$$

where $a_{\alpha,j} = \langle \boldsymbol{t}^{(\alpha)}, \boldsymbol{\Delta}_j^{(\alpha)}\rangle$. Secondly, define also

$$\breve{R}_{n,A}(\boldsymbol{t}) = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\prod_{k\in A}\left[\exp(\mathrm{i}\langle \boldsymbol{t}^{(k)}, \boldsymbol{\epsilon}_j^{(k)}\rangle) - \phi(\boldsymbol{t}^{(k)})\right]. \tag{6.7}$$

The proof proceeds with the following steps

$$\rho(\hat{R}_{n,A}, \tilde{R}_{n,A}) \xrightarrow{P} 0, \tag{6.8}$$

$$\rho(\tilde{R}_{n,A}, \breve{R}_{n,A}) \xrightarrow{P} 0, \tag{6.9}$$

$$\{\breve{R}_{n,A} :\ |A| > 1\} \Rightarrow \{R_A :\ |A| > 1\}. \tag{6.10}$$

The last step was proven in Theorem 2.1. A Taylor expansion of

$$\prod_{k\in A}\left[\exp(\mathrm{i}\langle \boldsymbol{t}^{(k)}, \boldsymbol{\epsilon}_j^{(k)}\rangle + a_{k,j}) - \phi(\boldsymbol{t}^{(k)})\right]$$

around $(a_{k,j})_{k\in A} = \mathbf{0}$ and Schwarz's inequality yield

$$\left|\hat{R}_{n,A}(\boldsymbol{t}) - \tilde{R}_{n,A}(\boldsymbol{t})\right| \leq 2^{|A|-2}|A|^2\|\boldsymbol{t}\|^2\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\|\boldsymbol{\Delta}_j\|^2, \tag{6.11}$$

where $\boldsymbol{\Delta}_j = (\boldsymbol{\Delta}_j^{(1)}, \ldots, \boldsymbol{\Delta}_j^{(p)})$. Now

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\|\boldsymbol{\Delta}_j\|^2 = \sum_{k=1}^{p}\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\|\boldsymbol{\Delta}_j^{(k)}\|^2, \tag{6.12}$$

where each of the $p$ terms is expressed as

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\|\boldsymbol{\Delta}_j^{(k)}\|^2 = \sqrt{n}\mathrm{tr}\left[(\boldsymbol{S}^{-\frac{1}{2}} - \boldsymbol{I})^2\frac{1}{n}\sum_{j=1}^{n}\boldsymbol{\epsilon}_j^{(k)}\boldsymbol{\epsilon}_j^{(k)\mathsf{T}}\right] \tag{6.13}$$

$$- 2\overline{\boldsymbol{\epsilon}}^{(k)\mathsf{T}}\boldsymbol{S}^{-\frac{1}{2}}\sqrt{n}(\boldsymbol{S}^{-\frac{1}{2}} - \boldsymbol{I})\overline{\boldsymbol{\epsilon}} + \sqrt{n}\overline{\boldsymbol{\epsilon}}^{\mathsf{T}}\boldsymbol{S}^{-1}\overline{\boldsymbol{\epsilon}}. \tag{6.14}$$

In view of $\overline{\boldsymbol{\epsilon}}^{(k)} = O_P(n^{-\frac{1}{2}})$, $\overline{\boldsymbol{\epsilon}} = O_P(n^{-\frac{1}{2}})$,

$$\sqrt{np}(\boldsymbol{S}^{-\frac{1}{2}} - \boldsymbol{I}) = -\frac{1}{2\sqrt{np}} \sum_{j=1}^{n} \sum_{k=1}^{p} \left( \boldsymbol{\epsilon}_j^{(k)} \boldsymbol{\epsilon}_j^{(k)\mathsf{T}} - \boldsymbol{I} \right) + O_P(n^{-\frac{1}{2}}) \tag{6.15}$$

and

$$\frac{1}{n} \sum_{j=1}^{n} \boldsymbol{\epsilon}_j^{(k)} \boldsymbol{\epsilon}_j^{(k)\mathsf{T}} = \boldsymbol{I} + O_P(n^{-\frac{1}{2}}) \tag{6.16}$$

(see Henze and Wagner (1997), p.9), we obtain $\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \|\boldsymbol{\Delta}_j\|^2 \xrightarrow{P} 0$. Using (6.11), $\rho_k(\hat{R}_{n,A}, \tilde{R}_{n,A}) \xrightarrow{P} 0$ and so (6.8) is proved. To establish (6.9) consider $\rho_k(\tilde{R}_{n,A}(\boldsymbol{t}), \breve{R}_{n,A}(\boldsymbol{t})) = \sup_{\|\boldsymbol{t}\| \le k} |U_n(\boldsymbol{t})|$, where

$$U_n(\boldsymbol{t}) = \sum_{\alpha \in A} \left\{ \boldsymbol{t}^{(\alpha)\mathsf{T}} \sqrt{n}(\boldsymbol{S}^{-\frac{1}{2}} - \boldsymbol{I}) \left[ A_{n,\alpha}(\boldsymbol{t}^{(A)}) - \overline{\boldsymbol{\epsilon}} B_{n,\alpha}(\boldsymbol{t}^{(A)}) \right] \right. \tag{6.17}$$

$$\left. - \quad \boldsymbol{t}^{(\alpha)\mathsf{T}} \sqrt{n} \overline{\boldsymbol{\epsilon}} B_{n,\alpha}(\boldsymbol{t}^{(A)}) \right\} \tag{6.18}$$

with

$$A_{n,\alpha}(\boldsymbol{t}^{(A)}) = \frac{1}{n} \sum_{j=1}^{n} \left\{ \boldsymbol{\epsilon}_j^{(\alpha)} \exp(\mathrm{i}\langle \boldsymbol{t}^{(\alpha)}, \boldsymbol{\epsilon}_j^{(\alpha)} \rangle) \prod_{k \in A; k \ne \alpha} \left[ \exp(\mathrm{i}\langle \boldsymbol{t}^{(k)}, \boldsymbol{\epsilon}_j^{(k)} \rangle) - \phi(\boldsymbol{t}^{(k)}) \right] \right\} \tag{6.19}$$

and

$$B_{n,\alpha}(\boldsymbol{t}^{(A)}) = \frac{1}{n} \sum_{j=1}^{n} \left\{ \exp(\mathrm{i}\langle \boldsymbol{t}^{(\alpha)}, \boldsymbol{\epsilon}_j^{(\alpha)} \rangle) \prod_{k \in A; k \ne \alpha} \left[ \exp(\mathrm{i}\langle \boldsymbol{t}^{(k)}, \boldsymbol{\epsilon}_j^{(k)} \rangle) - \phi(\boldsymbol{t}^{(k)}) \right] \right\}. \tag{6.20}$$

Note that both expressions $A_{n,\alpha}$ and $B_{n,\alpha}$ are i.i.d. averages of zero mean variables. By the uniform strong law of large numbers on compact sets (see (Ferguson, 1996, p. 108)), one can show that

$$\max_{\|\boldsymbol{t}^{(A)}\| \le k} \|A_{n,\alpha}(\boldsymbol{t}^{(A)})\| \xrightarrow{as} 0 \tag{6.21}$$

and

$$\max_{\|\boldsymbol{t}^{(A)}\| \le k} |B_{n,\alpha}(\boldsymbol{t}^{(A)})| \xrightarrow{as} 0, \tag{6.22}$$

so that (6.9) is proved.

### 6.3. Proof of Theorem 2.3

Let $x_{n,j} = \int_{\mathbb{R}_j^{pq}} f(\boldsymbol{y}_n(\boldsymbol{t})) \mathrm{d}G(\boldsymbol{t})$ and $x_j = \int_{\mathbb{R}_j^{pq}} f(\boldsymbol{y}(\boldsymbol{t})) \mathrm{d}G(\boldsymbol{t})$. Define the mapping $h_j : C(\mathbb{R}_j^{pq}, \mathbb{C})^2 \to \mathbb{R}$ by $h_j(\boldsymbol{y}) = \int_{\mathbb{R}_j^{pq}} f(\boldsymbol{y}(\boldsymbol{t})) \mathrm{d}G(\boldsymbol{t})$. Then, $x_{n,j} = h_j(r_j(\boldsymbol{y}_n))$ and $x_j = h_j(r_j(\boldsymbol{y}))$. By assumption, for all $j$, $P_{\boldsymbol{y}_n} r_j^{-1} \Rightarrow P_{\boldsymbol{y}} r_j^{-1}$. Thus, Theorem 5.1 of Billingsley (1968) and the continuity of $h_j$ imply $P_{x_{n,j}} \Rightarrow P_{x_j}$ as $n \to \infty$, for all $j$. Also, the dominated convergence theorem yields $x_j \xrightarrow{\mathcal{D}} w$, as $j \to \infty$. Using Theorem 4.2 of Billingsley (1968), in order to establish $w_n \xrightarrow{\mathcal{D}} w$ as $n \to \infty$, it suffices to show that for all $\epsilon > 0$,

$$\lim_{j \to \infty} \limsup_{n \to \infty} P\{|x_{n,j} - w_n| \ge \epsilon\} = 0.$$

By Markov's inequality it suffices to show that for all $\epsilon > 0$,

$$\lim_{j\to\infty} \limsup_{n\to\infty} E|x_{n,j} - w_n|^\alpha = 0 \tag{6.23}$$

for some $\alpha > 0$. But

$$\begin{aligned}
E|x_{n,j} - w_n|^\alpha &= E\left|\int_{\mathbb{R}^{pq}\setminus\mathbb{R}^{pq}_j} f(\boldsymbol{y}_n(\boldsymbol{t}))\mathrm{d}G(\boldsymbol{t})\right|^\alpha \\
&\leq E\left[\int_{\mathbb{R}^{pq}\setminus\mathbb{R}^{pq}_j} |f(\boldsymbol{y}_n(\boldsymbol{t}))|\mathrm{d}G(\boldsymbol{t})\right]^\alpha \\
&= E_{\boldsymbol{y}_n}\left(E_G\left[\mathbb{1}\{\|\boldsymbol{t}\| > j\}|f(\boldsymbol{y}_n(\boldsymbol{t}))|\right]\right)^\alpha .
\end{aligned}$$

Then, using Jensen's inequality with $\alpha \geq 1$ and Fubini's theorem,

$$\begin{aligned}
E|x_{n,j} - w_n|^\alpha &\leq E_{\boldsymbol{y}_n} E_G[\mathbb{1}\{\|\boldsymbol{t}\| > j\}|f(\boldsymbol{y}_n(\boldsymbol{t}))|^\alpha] \\
&= E_G E_{\boldsymbol{y}_n}[\mathbb{1}\{\|\boldsymbol{t}\| > j\}|f(\boldsymbol{y}_n(\boldsymbol{t}))|^\alpha] \\
&= \int_{\mathbb{R}^{pq}\setminus\mathbb{R}^{pq}_j} E|f(\boldsymbol{y}_n(\boldsymbol{t}))|^\alpha \mathrm{d}G(\boldsymbol{t}).
\end{aligned}$$

The theorem is proved.

### 6.4. Proof of Theorem 2.4

Define the functional (norm) $|[x]|^2 = \int |x(\boldsymbol{t})|^2\varphi_b(\boldsymbol{t})\mathrm{d}\boldsymbol{t}$ on the subset of squared-integrable functions with respect to $\varphi_b(\boldsymbol{t})d\boldsymbol{t}$. Following (Henze and Wagner, 1997, pp. 10-11) , in order to prove the theorem, it suffices to show

$$\left(|[\breve{R}_{n,A}]|^2, |[\breve{R}_{n,B}]|^2\right) \xrightarrow{\mathcal{D}} \left(|[R_A]|^2, |[R_B]|^2\right) \tag{6.24}$$

and

$$|[\tilde{R}_{n,A} - \breve{R}_{n,A}]|^2 \xrightarrow{P} 0. \tag{6.25}$$

This is sufficient since then the triangle inequality produces

$$\left||[\breve{R}_{n,A}]| - |[\tilde{R}_{n,A}]|\right| \leq |[\breve{R}_{n,A} - \tilde{R}_{n,A}]| \xrightarrow{P} 0.$$

This implies

$$\left(|[\tilde{R}_{n,A}]|^2, |[\tilde{R}_{n,B}]|^2\right) \xrightarrow{\mathcal{D}} \left(|[R_A]|^2, |[R_B]|^2\right). \tag{6.26}$$

But from (6.11), $\left|\hat{R}_{n,A}(\boldsymbol{t}) - \tilde{R}_{n,A}(\boldsymbol{t})\right| = \|\boldsymbol{t}\|^2 o_P(1)$ which yields

$$|[\hat{R}_{n,A} - \tilde{R}_{n,A}]|^2 \xrightarrow{P} 0. \tag{6.27}$$

Again, the triangle inequality produces

$$\left||[\hat{R}_{n,A}]| - |[\tilde{R}_{n,A}]|\right| \leq |[\hat{R}_{n,A} - \tilde{R}_{n,A}]| \xrightarrow{P} 0.$$

Then, the desired result

$$\left(|[\hat{R}_{n,A}]|^2, |[\hat{R}_{n,B}]|^2\right) \xrightarrow{\mathcal{D}} \left(|[R_A]|^2, |[R_B]|^2\right) \tag{6.28}$$

could be concluded.

The convergence in (6.24) is now proved by means of arbitrary linear combinations with the use of Theorem 2.3 with $\boldsymbol{y}_n = (\breve{R}_{n,A}, \breve{R}_{n,B})$, and the continuous function $f : \mathbb{C}^2 \to \mathbb{R}$ defined by $f(x_1, x_2) = a_1|x_1|^2 + a_2|x_2|^2$ for arbitrary constants $a_1$ and $a_2$ . Theorem 2.1 states that $(\breve{R}_{n,A}, \breve{R}_{n,B}) \xrightarrow{\mathcal{D}} (R_A, R_B)$. From Proposition 14.6 of Kallenberg (1997), this remains true on all the closed balls. Note that

$$E|f(\boldsymbol{y}_n(\boldsymbol{t}))| \leq E|a_1||\breve{R}_{n,A}(\boldsymbol{t})|^2 + E|a_2||\breve{R}_{n,B}(\boldsymbol{t})|^2.$$

It can be readily shown that $E|\breve{R}_{n,A}(\boldsymbol{t})|^2 = \prod_{k\in A}(1 - \phi^2(\boldsymbol{t}^{(k)}))$, which is independent of $n$ and integrable with respect to $\varphi_b(\boldsymbol{t})d\boldsymbol{t}$. Hence, condition (2.13) is satisfied with $\alpha = 1$. The convergence in (6.24) thus holds.

To prove (6.25), proceed as follows. From (6.4), (6.7), (6.17) and the inequality $(\sum_{j=1}^{s} a_j)^2 \leq s\sum_{j=1}^{s} a_j^2$, it follows

$$|\tilde{R}_{n,A}(\boldsymbol{t}) - \breve{R}_{n,A}(\boldsymbol{t})|^2 \leq \tag{6.29}$$

$$|A|\sum_{\alpha\in A}\left[\frac{\|\boldsymbol{t}\|^2}{4}\left\|\frac{1}{\sqrt{np}}\sum_{j=1}^{n}\sum_{k=1}^{p}(\boldsymbol{\epsilon}_j^{(k)}\boldsymbol{\epsilon}_j^{(k)\intercal} - \boldsymbol{I})\right\|^2 \left\|A_{n,\alpha}(\boldsymbol{t}^{(A)})\right\|^2\right. \tag{6.30}$$

$$+\|\boldsymbol{t}\|^2\left\|A_{n,\alpha}(\boldsymbol{t}^{(A)})\right\|^2 o_P(1) + \|\boldsymbol{t}\|^2 n\left\|\overline{\boldsymbol{\epsilon}}\right\|^2\left|B_{n,\alpha}(\boldsymbol{t}^{(A)})\right|^2 \tag{6.31}$$

$$\left.+\|\boldsymbol{t}\|^2\left\|\sqrt{n}(\boldsymbol{S}^{-\frac{1}{2}} - \boldsymbol{I})\right\|^2\left\|\overline{\boldsymbol{\epsilon}}\right\|^2\left|B_{n,\alpha}(\boldsymbol{t}^{(A)})\right|^2\right]. \tag{6.32}$$

Let

$$\begin{aligned} W_n &= \int_{\mathbb{R}^p}\left|B_{n,\alpha}(\boldsymbol{t}^{(A)})\right|^2\|\boldsymbol{t}\|^2\varphi_b(\boldsymbol{t})\mathrm{d}\boldsymbol{t}, \\ \tilde{W}_n &= \int_{\mathbb{R}^p}\left\|A_{n,\alpha}(\boldsymbol{t}^{(A)})\right\|^2\|\boldsymbol{t}\|^2\varphi_b(\boldsymbol{t})\mathrm{d}\boldsymbol{t}. \end{aligned}$$

By Tonelli's theorem,

$$EW_n = \int E\left|B_{n,\alpha}(\boldsymbol{t}^{(A)})\right|^2\|\boldsymbol{t}\|^2\varphi_b(\boldsymbol{t})\mathrm{d}\boldsymbol{t}, \tag{6.33}$$

where by direct calculation

$$E\left|B_{n,\alpha}(\boldsymbol{t}^{(A)})\right|^2 = \frac{1}{n}\prod_{k\in A;k\neq\alpha}(1 - \phi^2(\boldsymbol{t}^{(k)})), \tag{6.34}$$

so that $W_n = o_P(1)$. Similarly, since

$$E\left\|A_{n,\alpha}(\boldsymbol{t}^{(A)})\right\|^2 = \frac{p}{n}\prod_{k\in A;k\neq\alpha}(1 - \phi^2(\boldsymbol{t}^{(k)})), \tag{6.35}$$

it follows that $\tilde{W}_n = o_P(1)$. Thus,

$$|[\tilde{R}_{n,A} - \breve{R}_{n,A}]|^2 \leq o_P(1). \tag{6.36}$$

### 6.5. Proof of Theorem 3.1

As in the proof of Theorem 2.1, it suffices to show that $r_j(R_{n,A}) \xrightarrow{fd} r_j(R_A)$ and $\{r_j(R_{n,A})\}$ is a tight family. We have

$$\begin{aligned} R_{n,A}(\boldsymbol{t}) &= \frac{1}{\sqrt{n}} \sum_{l=1}^{pc_n} \prod_{k \in A} \left[ \exp\left( \mathrm{i} \langle \boldsymbol{t}^{(k)}, \boldsymbol{\epsilon}_l^{(k)} \rangle \right) - \phi(\boldsymbol{t}^{(k)}) \right] \\ &\quad + \frac{1}{\sqrt{n}} \sum_{l=pc_n+1}^{n-p+1} \prod_{k \in A} \left[ \exp\left( \mathrm{i} \langle \boldsymbol{t}^{(k)}, \boldsymbol{\epsilon}_l^{(k)} \rangle \right) - \phi(\boldsymbol{t}^{(k)}) \right] \\ &\equiv \tilde{R}_{n,A}(\boldsymbol{t}) + r_n(\boldsymbol{t}), \end{aligned}$$

where $c_n = \left[ \frac{n-p+1}{p} \right]$ is the integer part. The process $R_{n,A}$ is asymptotically equivalent to $\tilde{R}_{n,A}$ since

$$\rho_j(R_{n,A}, \tilde{R}_{n,A}) \leq \frac{1}{\sqrt{n}} \sum_{l=pc_n+1}^{n-p+1} 2^{|A|} \xrightarrow{P} 0. \tag{6.37}$$

Let

$$Y_l^A(\boldsymbol{t}) = \prod_{k \in A} \left[ \exp\left( \mathrm{i} \langle \boldsymbol{t}^{(k)}, \boldsymbol{\epsilon}_l^{(k)} \rangle \right) - \phi(\boldsymbol{t}^{(k)}) \right], \; l = 1, 2, \ldots, \tag{6.38}$$

which is an $m$-dependent sequence. Thus, using a multivariate central limit theorem for $m$-dependent complex random vectors obtained by applying the Cramér-Wold device in connection with a complex variant of Theorem 11 in Ferguson (1996), the finite dimensional convergence to Gaussian limit and the covariance function are obtained. It remains to show that $\{\tilde{R}_{n,A}\}$ is a tight sequence. Since $\{1, 2, \ldots, pc_n\} = \cup_{h=1}^{p} \{pl + h;\ 0 \leq l < c_n\}$, we have $r_j(\tilde{R}_{n,A})(\boldsymbol{t}) = \sum_{h=1}^{p} \sum_{B \subset A} (-1)^{|A \setminus B|} \phi_{A,B}(\boldsymbol{t}) Y_{n,h,B}^j(\boldsymbol{t})$, where $\phi_{A,B}(\boldsymbol{t}) = \prod_{l \in A \setminus B} \phi(\boldsymbol{t}^{(l)})$ and

$$Y_{n,h,B}^j(\boldsymbol{t}) = \frac{1}{\sqrt{n}} \sum_{l=0}^{c_n - 1} \left[ \prod_{j \in B} \exp\left( \mathrm{i} \langle \boldsymbol{t}^{(j)}, \boldsymbol{\epsilon}_{pl+h}^{(j)} \rangle \right) - \prod_{j \in B} \phi(\boldsymbol{t}^{(j)}) \right]. \tag{6.39}$$

To show that $\{Y_{n,h,B}^j\}$ is tight, proceed as in proof of Theorem 2.1. Since there is only a finite number of $h$'s and $B$'s, tightness follows and the theorem is proved.

### 6.6. Proof of Theorem 3.2

Let

$$\tilde{R}_{n,A}(\boldsymbol{t}) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n-p+1} \prod_{k \in A} [\exp(\mathrm{i} \langle \boldsymbol{t}^{(k)}, \boldsymbol{\epsilon}_j^{(k)} \rangle) - \phi(\boldsymbol{t}^{(k)})] + \mathrm{i} U_n(\boldsymbol{t}), \tag{6.40}$$

where $U_n(\boldsymbol{t})$ is defined as in (6.4) and where $\boldsymbol{\Delta}_j^{(k)} = (\boldsymbol{S}^{-\frac{1}{2}} - \boldsymbol{I}) \boldsymbol{\epsilon}_j^{(k)} - \boldsymbol{S}^{-\frac{1}{2}} \bar{\boldsymbol{u}}$. Let also

$$\breve{R}_{n,A}(\boldsymbol{t}) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n-p+1} \prod_{k \in A} \left[ \exp(\mathrm{i} \langle \boldsymbol{t}^{(k)}, \boldsymbol{\epsilon}_j^{(k)} \rangle) - \phi(\boldsymbol{t}^{(k)}) \right]. \tag{6.41}$$

Proceed as in the proof of Theorem 2.2 to establish (6.8), (6.9) and (6.10). The proofs of (6.8) and (6.9) are established along the lines in the proof of Theorem 2.2. Step (6.10) was proven in Theorem 3.1.

### 6.7. Proof of Theorem 3.3

Follow the lines in the proof of Theorem 2.4 but use Theorem 3.1 instead of Theorem 2.1, and replace (6.29) with

$$|\tilde{R}_{n,A}(\boldsymbol{t}) - \breve{R}_{n,A}(\boldsymbol{t})|^2 \leq \tag{6.42}$$

$$|A| \sum_{\alpha \in A} \left[ \frac{\|\boldsymbol{t}\|^2}{4} \left\| \frac{1}{\sqrt{n}} \sum_{j=1}^{n} (\boldsymbol{u}_j \boldsymbol{u}_j^{\mathsf{T}} - \boldsymbol{I}) \right\|^2 \left\| A_{n,\alpha}(\boldsymbol{t}^{(A)}) \right\|^2 \right. \tag{6.43}$$

$$+ \|\boldsymbol{t}\|^2 \left\| A_{n,\alpha}(\boldsymbol{t}^{(A)}) \right\|^2 o_P(1) + \|\boldsymbol{t}\|^2 n \left\| \overline{\boldsymbol{u}} \right\|^2 \left| B_{n,\alpha}(\boldsymbol{t}^{(A)}) \right|^2 \tag{6.44}$$

$$\left. + \|\boldsymbol{t}\|^2 \left\| \sqrt{n}(\boldsymbol{S}^{-\frac{1}{2}} - \boldsymbol{I}) \right\|^2 \left\| \overline{\boldsymbol{u}} \right\|^2 \left| B_{n,\alpha}(\boldsymbol{t}^{(A)}) \right|^2 \right]. \tag{6.45}$$

Replace also $n$ with $n - p + 1$ in (6.34) and (6.35).

## ACKNOWLEDGEMENTS

The first author would like to thank the National Sciences and Engineering Research Council of Canada for financial support through grant 97303-99. This paper is part of the second author's Ph.D. thesis.

## BIBLIOGRAPHY

[1] Baringhaus, L., Henze, N., 1988. A consistent test for multivariate normality based on the empirical characteristic function. Metrika 35 (6), 339–348.
[2] Billingsley, P., 1968. Convergence of probability measures. John Wiley & Sons Inc., New York.
[3] Blum, J. R., Kiefer, J., Rosenblatt, M., 1961. Distribution free tests of independence based on the sample distribution function. Ann. Math. Statist. 32, 485–498.
[4] Chen, G., Lockhart, R. A., Stephens, M. A., 2002. Box-cox transformations in linear models: Large sample theory and tests of normality. Canad. J. Statist. 30 (2), 177–234.
[5] Conway, J. B., 1985. A course in functional analysis. Springer-Verlag, New York.
[6] Cools, R., 1999. Monomial cubature rules since "Stroud": a compilation. II. J. Comput. Appl. Math. 112 (1-2), 21–27, numerical evaluation of integrals.
[7] Cools, R., Rabinowitz, P., 1993. Monomial cubature rules since "Stroud": a compilation. J. Comput. Appl. Math. 48 (3), 309–326.
[8] Csörgő, S., 1981a. Limit behaviour of the empirical characteristic function. Ann. Probab. 9 (1), 130–144.
[9] Csörgő, S., 1981b. Multivariate empirical characteristic functions. Z. Wahrsch. Verw. Gebiete 55 (2), 203–229.
[10] Csörgő, S., 1985. Testing for independence by the empirical characteristic function. J. Multivariate Anal. 16 (3), 290–299.

[11] Davies, R. B., 1973. Numerical inversion of a characteristic function. Biometrika 60, 415–417.

[12] Davies, R. B., 1980. [Algorithm AS 155] The distribution of a linear combination of $\chi^2$ random variables (AS R53: 84V33 p366- 369). Appl. Statistics 29, 323–333.

[13] de Wet, T., Randles, R. H., 1987. On the effect of substituting parameter estimators in limiting $\chi^2$ $U$ and $V$ statistics. Ann. Statist. 15 (1), 398–412.

[14] Deheuvels, P., Martynov, G. V., 1996. Cramér-von Mises-type tests with applications to tests of independence for multivariate extreme-value distributions. Comm. Statist. Theory Methods 25 (4), 871–908.

[15] Ferguson, T. S., 1996. A course in large sample theory. Chapman & Hall, London.

[16] Feuerverger, A., 1993. A consistent test for bivariate dependence. Internat. Statist. Rev. 61 (2), 419–433.

[17] Feuerverger, A., Mureika, R. A., 1977. The empirical characteristic function and its applications. Ann. Statist. 5 (1), 88–97.

[18] Ghoudi, K., Kulperger, R. J., Rémillard, B., 2001. A nonparametric test of serial independence for time series and residuals. J. Multivariate Anal. 79, 191–218.

[19] Henze, N., Wagner, T., 1997. A new approach to the BHEP tests for multivariate normality. J. Multivariate Anal. 62, 1–23.

[20] Henze, N., Zirkler, B., 1990. A class of invariant consistent tests for multivariate normality. Comm. Statist. Theory Methods 19, 3595–3617.

[21] Imhof, J. P., 1961. Computing the distribution of quadratic forms in normal variables. Biometrika 48, 419–426.

[22] Kallenberg, O., 1997. Foundations of modern probability. Springer-Verlag, New York.

[23] Kellermeier, J., 1980. The empirical characteristic function and large sample hypothesis testing. J. Multivariate Anal. 10 (1), 78–87.

[24] Lee, Y.-S., Lee, M. C., 1992. On the derivation and computation of the Cornish-Fisher expansion. Austral. J. Statist. 34 (3), 443–450.

[25] Lee, Y.-S., Lin, T.-K., 1992. [Algorithm AS 269] High order Cornish-Fisher expansion. Appl. Statistics 41, 233–240.

[26] Marcus, M. B., 1981. Weak convergence of the empirical characteristic function. Ann. Probab. 9 (2), 194–201.

[27] Rencher, A. C., 2002. Methods of multivariate analysis, 2nd Edition. Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], New York.

[28] Stroud, A., 1967. Some seventh degree integration formulas for symmetric regions. SIAM J. Numer. Anal. 4, 37–44.

[29] Stroud, A. H., 1971. Approximate calculation of multiple integrals. Prentice-Hall Inc., Englewood Cliffs, N.J., prentice-Hall Series in Automatic Computation.

# CHAPITRE 4

# Conclusion

La question initiale de recherche consistait à explorer différentes pistes pouvant permettre de valider certaines hypothèses qui sont souvent faites dans le contexte de séries stochastiques dépendantes, et en particulier l'hypothèse de normalité et d'indépendance du bruit blanc pour lequel peu de travaux ont été publiés. Dans le Chapitre 2, nous nous sommes attaqués à la première partie de ce problème en considérant la mise au point d'un test de normalité pour les innovations d'un modèle ARMA univarié de moyenne connue. Les modèles ARMA univariés sont en effet très connus et utilisés dans la pratique. Cette construction utilise la stratégie des tests lisses de Neyman (1937) enrichie par la technologie « data-driven » de Ledwina (1994) qui sont, selon les travaux de Ledwina et son équipe, théoriquement plus performants que les autres approches. Le test obtenu autorise une bonne puissance en comparaison des rares tests existants et son niveau est atteint et tenu assez rapidement. Une extension de ce travail est en cours pour considérer le cas où la moyenne est inconnue, estimée par la méthode du maximum de vraisemblance. Il sera ensuite possible de généraliser les résultats obtenus au cas d'un ARMA multivarié, et probablement aussi à celui d'un ARIMA multivarié.

Le chapitre suivant aborde le problème consistant à tester l'indépendance de données sérielles. L'approche des tests lisses n'étant pas encore développée pour le cas de séries multivariées, des résultats théoriques sur la convergence faible de processus stochastiques à valeurs dans un espace de Fréchet ont été employés. En utilisant la notion de convergence des lois de dimension finie et la tension d'une famille, la convergence faible d'un certain processus empirique basé sur la fonction caractéristique a été démontrée. L'application d'une fonctionnelle de type Cramér-von Mises a alors permis l'obtention d'un test d'indépendance multivarié pour des marginales Gaussiennes fixées. Ce résultat est d'un intérêt non négligeable dans le domaine des données familiales. Il a ensuite été étendu au cas de données dépendantes afin de produire un test multivarié d'indépendance sérielle, qui pourrait possiblement être appliqué pour vérifier l'hypothèse de bruit blanc dans le cas de la régression multivariée et celui d'innovations Gaussiennes d'un modèle ARMA. En considérant le cas d'ensembles $A$ de cardinal unité, il sera possible de bâtir un test de l'hypothèse composée simultanément de la multinormalité et de l'indépendance. Pour cela, certains problèmes doivent encore être résolus comme la caractérisation de la loi de deux statistiques de Cramér-von

Mises dépendantes à partir de la donnée de leur covariance. Une autre extension importante consiste à se débarasser de l'hypothèse de multinormalité des marginales afin d'obtenir un test d'indépendance complètement non-paramétrique. Un théorème de Csörgő (1981) de type Glivenko-Cantelli pour la fonction caractéristique donne espoir d'atteindre cet objectif. Des développements récents sur les tests de normalité basés sur un processus, tels ceux exposés dans Chen et al. (2002), montrent l'intérêt actuel de nombreux chercheurs dans ce domaine.

## BIBLIOGRAPHIE

[1] Chen, G., Lockhart, R. A., Stephens, M. A., 2002. Box-cox transformations in linear models : Large sample theory and tests of normality. Canad. J. Statist. 30 (2), 177–234.

[2] Csörgő, S., 1981. Multivariate empirical characteristic functions. Z. Wahrsch. Verw. Gebiete 55 (2), 203–229.

[3] Ledwina, T., 1994. Data-driven version of Neyman's smooth test of fit. J. Amer. Statist. Assoc. 89 (427), 1000–1005.

[4] Neyman, J., 1937. Smooth test for goodness of fit. Skand. Aktuar. 20, 149–199.

# INDEX DES AUTEURS

# ANNEXE A

# Les Programmes Fortran 77 du premier article

Vu l'ampleur des programmes (plus de 13000 lignes de code !), le listing des différents programmes n'est fourni que dans la version électronique de ce document sur le site Internet `http://www.theses.umontreal.ca`.
Tous ces programmes ont été utilisés sur un PC (gracieusement mis à disposition par Gilles Ducharme) disposant d'un processeur de 1 mégaHz et pourvu, par mes soins, du système d'exploitation Linux Redhat 7.2.

Mode d'utilisation de mon programme de simulation :

Il faut commencer par changer les valeurs des paramètres souhaitées dans les programmes *big_prog_AR1.f*, *big_prog_AR2.f*, *big_prog_MA1.f*, *big_prog_MA2.f*, *big_prog_ARMA11.f*, *big_prog_ARMA12.f*, *big_prog_ARMA21.f*, *big_prog_ARMA22.f*. Les paramètres où des changements peuvent être nécessaires sont marqués, dans le code source, par le signe **.

Ensuite, il faut remplir le fichier nommé *paramentree* de la façon suivante : sur chaque ligne du fichier, il doit figurer en séquence les valeurs de $p$ $q$ $loi$ $\boldsymbol{\varphi}_1$ $\boldsymbol{\varphi}_2$ $\boldsymbol{\theta}_1$ $\boldsymbol{\theta}_2$ avec $p, q = 0, 1$ ou 2, $loi \in \{0, 1, 2, 3, 4, ..., 19\}$ et $\boldsymbol{\varphi}$ et $\boldsymbol{\theta}$ dans la région assurant l'inversibilité et la causalité du processus.

**Remarque 1.** *Même si $p$ (resp. $q$) =0, il faut donner des valeurs aux $\boldsymbol{\theta}$ (resp. aux $\boldsymbol{\varphi}$), mais ces valeurs ne seront pas prises en compte par le programme.*

Certains de ces fichiers, qui correspondent aux points de la grille que l'on a choisi avec Gilles Ducharme pour être dans les régions d'inversibilité, sont déjà créés dans les répertoires *FICHIERS_paramentree*. Il suffit de les mettre dans le dossier principal et de les renommer en *paramentree*.

Ensuite, on tape :
```
./compile
```

Puis :
```
./a.out
```

Les résultats sont alors stockés dans le dossier :
```
./SIMUL/
```
sous la forme de différents fichiers *para.i* et *resultat.i.xc*. Les fichiers *para.i* contiennent différents paramètres de la simulation. Les fichiers *resultat.i.xc* contiennent de nombreux résultats de la simulation et sont faciles à lire avec le logiciel Excel. Il est aussi créé un fichier *tableau.txt* qui résume cette information.

**Attention**
Si la simulation est interrompue, des fichiers qui sont normalement effacés à la fin de chaque simulation (*data.txt* et *ftemp*) vont subsister dans le dossier
```
./SIMUL/../.
```
Avant de relancer une autre série de simulations, il faut effacer ces deux fichiers. Il faut aussi déplacer (ou effacer si on ne veut plus les conserver) les fichiers *para.i*, *resultat.i.xc* et *tableau.txt* du dossier
```
SIMUL.
```

Déroulement des divers programmes :

Voilà comment se déroule l'exécution de mes programmes.
Le programme principal est le programme *test.f* (qui sera compilé en *a.out*).

1) Dans le programme **TEST** :
On crée le fichier *tableau.txt* qui contiendra sur chaque ligne :
Nbcle $T\ p\ q\ loi\ \alpha\ \boldsymbol{\varphi}_1\ \boldsymbol{\varphi}_2\ \boldsymbol{\theta}_1\ \boldsymbol{\theta}_2\ PuissR1\ PuissR2\ PuissR3\ PuissR4\ PuissR5\ PuissLed1\ PuissLed2\ PuissBD\ PuissAD\ PuissJB$.
Le programme **TEST** lit, dans le fichier *paramentree*, les valeurs de $p$, $q$, $loi$, $phi(1)$, $phi(2)$, $theta(1)$, $theta(2)$ et les stocke dans la matrice $temp$, cela permet aussi de calculer le nombre de lignes (*nbline*) du fichier *paramentree*.
Pour $i = 1$ à $nbline$ :
On crée le fichier *ftemp* qui contient la ligne $i$ du fichier *paramentree* en ne gardant bien sûr que ce qui est nécessaire. Ensuite on appelle, suivant les valeurs dans le fichier *ftemp*, le bon programme **.out* (–>2) puis on efface *ftemp* et on recommence à $i + 1$.
À chaque boucle, le programme **TEST** appelle donc l'un des programmes **ARMA22.out**, **ARMA21.out**, A**RMA12.out**, **ARMA11.out**, **MA2.out**, **MA1.out**, **AR2.out**, **AR1.out** ou **ARMA00.out**.

2) Prenons l'exemple où on appelle **ARMA22.out** :
Dans **ARMA22.out**, on va lire les paramètres nécessaires dans *ftemp* : $j$, $loi$, $phi(1)$, $phi(2)$, $theta(1)$, $theta(2)$.
On appelle le programme **creerdat_ARMA** (–>3) qui crée le fichier de données *data.txt* qui contient sur chaque ligne $sigchap2$ et $epschap$.
On appelle le programme **calcstat** (–>4) qui crée le fichier des statistiques *resultat.i.xc* (qui contient sur chaque ligne $numero - simul\ R1\ R2\ R3\ R4\ R5\ R6\ R7\ R8\ R9\ R10\ Kchap1\ Ledwi1\ Kchap2\ Ledwi2\ BD\ AD\ JB$) et rajoute au fichier *tableau.txt* une ligne de puissances.

On efface *data.txt*.
On sort et on se retrouve donc dans la boucle du programme **TEST**.

3) Dans le programme **creerdat_ARMA** :
On simule les données :

* si $loi = 0$ :
  On appelle **G05EGF** qui "sets up a reference vector for an ARMA model with normally distributed errors. It also initialises the series to a stationary position."
  On appelle **G05EWF** qui "returns the next term from an ARMA time series using the reference vector set up by **G05EGF**".

* si $loi \neq 0$ :
  On appelle **simARM** (–>5) (resp. **simAR**, resp. **simMA**) Les données seront les $Mind+p+1$ à $Mind+p+nT$ valeurs de $YtpMn$ (resp. $YtpMn$, resp. $YtMn$).

Puis, on estime les paramètres : appel de **G13DCF**.
Enfin, on crée le fichier *param.i* des paramètres de la simulation et on sort pour se retrouver dans **ARMA22.out**, qui va ensuite nous envoyer dans **calcstat**.

4) Dans **calcstat** :
On met les données du fichier *data.txt* dans la matrice $valeurs$.
On crée le vecteur des $sigchap2$.
On crée la matrice des $epschap$.
On commence la grande boucle de 1 à 10000.
On affiche un décompte à l'écran.
On calcule toutes les statistiques que l'on compare aux quantiles appropriés pour incrémenter (ou pas) le compteur qui permettra de calculer les puissances.
Fin de la grande boucle.

On calcule la puissance des tests en % en se servant des compteurs créés.
On calcule les quantiles.
On calcule les rangs des valeurs des statistiques.
On calcule les moyennes et écart-types des valeurs des statistiques.
On calcule le nombre de fois où chaque polynôme a été choisi par les procédures de Ledwina.
On écrit les résultats dans *resultat.ijkl.xc*.
On écrit quelques résultats dans *para.ijkl*.
On écrit les résultats dans *tableau.txt*.
On sort pour se retrouver dans **ARMA22.out**.

5) Dans **simulARMA** :

Ce programme réalise les étapes 2 et 3 de la partie « Simulation Algorithm 1 » de l'article de Burn. Cette sous-routine remplace le vecteur $YtpMn$ en entrée par $p+Mind+n$ données simulées d'un modèle ARMA(p,q) de vecteur d'innovations de loi donnée par l'entier $loi$ , les $p$ premières valeurs sont les $p$ valeurs initiales de l'étape 1 de Burn,

les $Mind$ valeurs suivantes sont les valeurs à écarter de la warm-up period de longueur $Mind$. Par conséquent, seules les $n$ dernières valeurs devront être conservées pour la suite de la simulation.
Elle renvoie aussi le vecteur $shocks$ des $marret + p$ innovations utilisées dans l'algorithme « Initialisation 1 » de Burn, ainsi que le vecteur $Atnq$ des $n + Mind + q$ « random shocks » $At$ de l'étape 2 de Burn.

```
c Calcul des valeurs de psi (sauf pour un MA)
      CALL ARMpsi

c Calcul de shocks et Wtip (sauf pour un MA)
      CALL shock

c Étape 2 de l'algorithme: on génère les random shocks.
c On appelle le bon générateur aléatoire suivant la valeur
c  de loi (loi=1,...,19).
Le vecteur Atnq est donc créé.

c Étape 3 de l'algorithme
c YtpMn:  Y(1-p), Y(2-p), ..., Y(0), Y(1) , ..., Y(Mind+n)
c YtpMn(i),i: 1     2           p    p+1         p+Mind+n
c Wtp:          Y(1-p), Y(2-p), ..., Y(0)
c YtpMn(i),i:    1       2           p
c Atnq: eps(1-q),eps(2-q),...,eps(0),eps(1),...,eps(n+Mind)
c Atnq(k), k: 1     2            q     q+1       n+Mind+q
```

Le vecteur $YtpMn$ est alors créé.

## A.1. LE SCRIPT *compile*

```
fort77 -c -f test.f
fort77 -c big_prog_ARMA22.f
fort77 -c big_prog_ARMA21.f
fort77 -c big_prog_ARMA12.f
fort77 -c big_prog_ARMA11.f
fort77 -c big_prog_AR2.f
fort77 -c big_prog_AR1.f
fort77 -c big_prog_MA2.f
fort77 -c big_prog_MA1.f
fort77 -c big_prog_ARMA00.f
f77 test.o -lnag
f77 -o ARMA22.out big_prog_ARMA22.o -lnag
f77 -o ARMA21.out big_prog_ARMA21.o -lnag
f77 -o ARMA12.out big_prog_ARMA12.o -lnag
f77 -o ARMA11.out big_prog_ARMA11.o -lnag
f77 -o AR2.out big_prog_AR2.o -lnag
f77 -o AR1.out big_prog_AR1.o -lnag
f77 -o MA2.out big_prog_MA2.o -lnag
f77 -o MA1.out big_prog_MA1.o -lnag
f77 -o ARMA00.out big_prog_ARMA00.o -lnag
rm -f *.o
```

## A.2. LISTE DES DIFFÉRENTS PROGRAMMES

```
Polynômes de Legendre:
-->H1.f
-->H2.f
-->H3.f
-->H4.f
-->H5.f
-->H6.f
-->H7.f
-->H8.f
-->H9.f
-->H10.f
Polynômes de Legendre modifiés (sans les ak):
-->H1etoile.f
-->H2etoile.f
-->H3etoile.f
-->H4etoile.f
-->H5etoile.f
-->H6etoile.f
-->H7etoile.f
-->H8etoile.f
-->H9etoile.f
-->H10etoile.f
Polynômes de Legendre modifiés (avec les ak):
-->H1isa.f
-->H2isa.f
-->H3isa.f
-->H4isa.f
-->H5isa.f
-->H6isa.f
-->H7isa.f
-->H8isa.f
-->H9isa.f
-->H10isa.f
Calcul des psi(i) comme dans l'article de Burn:
-->ARpsi.f
-->MApsi.f
-->ARMpsi.f
Calcule la densité d'un échantillon pour une loi de Laplace:
-->dlap.f
Calcule la densité d'un échantillon pour une loi Normale:
-->dnorm.f
Calcule la densité d'un échantillon pour une loi Skew-Normale:
-->dskew.f
Calcule le maximum d'un vecteur:
-->max.f
Calcule la moyenne d'un vecteur:
-->mean.f
Calcule le minimum d'un vecteur:
-->min.f
Calcule les probabilités d'une loi Normale(mu,sigma) aux points d'un  vecteur:
-->pnorm.f
Calcule les quantiles d'une loi Normale(mu,sigma^2):
-->qnorm.f
Simulation d'un échantillon d'une loi de Laplace:
-->rlap.f
Simulation d'un échantillon d'une loi Skew-Normale:
-->rlap.f
Ce programme réalise la partie Initialisation Algoritm 1, de l'article de Burn
Cette sous-routine remplace le vecteur Wtip en entrée par p données
simulées
d'un modèle ARMA(p,q)
de vecteur d'innovations de loi donnée par l'entier loi
elle renvoie aussi le vecteur shocks des marret+p innovations utilisées:
```

```
-->shock.f
Calcule la variance d'un vecteur:
-->var.f
Ce programme réalise les étapes 2 et 3 de Simulation Algorithm 1, de l'article de Burn
Cette sous-routine remplace le vecteur YtpMn en entrée par p+M+n données simulées
d'un modèle AR(p) de vecteur d'innovations de loi donnée par l'entier loi ,
les p premières valeurs sont les p valeurs initiales de l'étape 1 de Burn,
les M valeurs suivantes sont les valeurs à écarter de la warm-up
period de
longueur Mind
donc seules les n dernières valeurs devront être conservées pour la
suite de la
simulation
elle renvoie aussi le vecteur shocks des marret+p innovations
utilisées dans
l'algorithme
Initialisation 1 de Burn, ainsi que le vecteur Atn des n+Mind random
shocks At de
l'étape 2 de Burn.
-->simulAR.f
Ce programme réalise les étapes 2 et 3 de Simulation Algorithm 1, de l'article de Burn
Cette sous-routine remplace le vecteur YtpMn en entrée par p+Mind+n données simulées
d'un modèle ARMA(p,q) de vecteur d'innovations de loi donnée par l'entier loi ,
les p premières valeurs sont les p valeurs initiales de l'étape 1 de Burn,
les Mind valeurs suivantes sont les valeurs à écarter de la warm-up period de
longueur Mind
donc seules les n dernières valeurs devront être conservées pour la
suite de la simulation
elle renvoie aussi le vecteur shocks des marret+p innovations
utilisées dans l'algorithme
Initialisation 1 de Burn, ainsi que le vecteur Atnq des n+Mind+q
random shocks At de l'étape 2 de Burn.
-->simulARMA.f
Ce programme réalise les étapes 2 et 3 de Simulation Algorithm 1, de l'article de Burn
Cette sous-routine remplace le vecteur YtMn en entrée par M+n données simulées
d'un modèle MA(q) de vecteur d'innovations de loi donnée par l'entier loi ,
les M valeurs suivantes sont les valeurs a écarter de la warm-up
period de longueur Mind
donc seules les n dernières valeurs devront être conservées pour la
suite de la simulation
elle renvoie aussi le vecteur Atnq des n+Mind+q random shocks At de l'étape 2 de Burn.
-->simulMA.f
Idée de ce  programme:
Étape1)
On crée un fichier de données 'data.txt' en utilisant certaines valeurs de paramètres
à lire dans un fichier d'entrée 'ftemp' crée par le programme test.f à l'aide du
fichier 'paramentree', on stocke les résultats autres que
les données dans un fichier 'param.i'
Étape 2)
ensuite on utilise le fichier 'data.txt'
avec le programme calcstat pour créer le fichier des statistiques 'resultat.i'
Étape 3)
Ensuite, on efface le fichier 'data.txt'
-->big_prog_AR1.f
-->big_prog_AR2.f
-->big_prog_MA1.f
-->big_prog_MA2.f
-->big_prog_ARMA11.f
-->big_prog_ARMA12.f
-->big_prog_ARMA21.f
-->big_prog_ARMA22.f
Ce programme crée le fichier de données data.txt qui contient, en ligne:
sigch2 et epschap:
-->creerdat_ARMA.f
-->creerdat_MA.f
```

```
-->creerdat_AR.f
Idée:
On crée un fichier ARMA22.out, ARMA21.out, ARMA12.out, ARMA11.out,
AR2.out, AR1.out, MA2.out et MA1.out qui vont lire des valeurs de p,q,phi,teta et loi
dans le fichier param_entree(c'est big_prog_ARMA.f, big_prog_AR.f,
big_prog_MA.f compilés pour
différentes valeurs de p et q)
Par exemple, si c'est MA2.out il laissera tomber la valeur de p et phi.
Ensuite, ce programme va appeler ces différents fichier.out successivement,
à l'aide de la commande 'shell'
-->test.f
```

# Annexe B

# Les Programmes C++ du deuxième article

Vu l'ampleur des programmes (presque 4500 lignes de code), le listing des différents programmes n'est fourni que dans la version électronique de ce document sur le site Internet `http://www.theses.umontreal.ca`.
Tous ces programmes ont été utilisés sur un PC (gracieusement mis à disposition par Gilles Ducharme) disposant d'un processeur de 1 mégaHz et pourvu, par mes soins, du système d'exploitation Linux Redhat 7.2.

Ces programmes utilisent les librairies *newmat10* et *newran02* programmées par Robert Davies et téléchargeables (fichiers *newmat10.tar.gz* et *newran02.tar.gz*) sur son site Internet via l'URL `http://www.robertnz.net/index.html`.

Pour compiler les programmes, il faut placer tous les fichiers *.c dans un même répertoire avec les fichiers *libnewmat.a* et *libnewran.a* (obtenus à partir des fichiers source mentionnés ci-dessus), et utiliser la commande :

```
g++ -Wall -O main.c -L. -lnewmat -lnewran -o main
```

Un fichier nommé *main* est alors créé, et c'est ce fichier qu'il faut utiliser pour le calcul des quantiles obtenus par la méthode de Imhof (ou Davies, ou Deheuvels et Martynov) ainsi que pour le calcul des valeurs prises par la fonction de répartition de la loi limite de la statistique de Cramér-von Mises :

```
./main
```

Pour obtenir les quantiles empiriques dans le cas non sériel, il faut pour compiler utiliser la commande :

```
g++ -Wall -O test.c -L. -lnewmat -lnewran -o test
```

et pour lancer la simulation de Monte-Carlo, la commande :

```
./test > resultat.txt
```

Lorsque la simulation est terminée, les résultats se trouvent dans le fichier *resultat.txt*.

Un autre fichier nommé *manova.c* contient le programme permettant d'obtenir les résultats de la section 5 du deuxième article.
Il faut à nouveau utiliser les commandes

```
g++ -Wall -O manova.c -L. -lnewmat -lnewran -o manova
```

et

```
./manova
```

On obtient ainsi les puissances de notre test et de celui de Wilks dans le cadre considéré.

# CURRICULUM VITAE

---

### Formation académique

- 1998-2002 PhD en statistique - Doctorat de Mathématiques Appliquées.
  Cotutelle France-Québec.
  Tests d'indépendance en analyse multivariée et tests de normalité dans les modèles ARMA.
  Directeurs de thèse : G. Ducharme et M. Bilodeau.
  Université Montpellier II et Université de Montréal.

- 1997-1998 D.E.A. de Biostatistique
  Equivalent à la 2ème année de maîtrise au Québec.
  Université Montpellier II.

- 1996-1997 Maîtrise d'Ingénierie Mathématique.
  Equivalent à la 1ère année de maîtrise au Québec.
  Université Montpellier II.

- 1994-1996 Licence de Mathématiques.
  Equivalent à la 3ème année de Baccalauréat au Québec.
  Université Montpellier II.

- 1992-1994 Deug A MP' Physique.
  Equivalent aux deux premières années de Baccalauréat au Québec.
  Université Montpellier II.

- 1991-1992 Première année de Mathématiques Supérieures.
  Lycée Joffre, Montpellier.

- 1991 Baccalauréat français série C.

### Coordonnées

- Email 1 : lafaye@dms.umontreal.ca
- Email 2 : plafaye@club-internet.fr
- Site Internet 1 : `http ://www.dms.umontreal.ca/~lafaye`
- Site Internet 2 : `http ://perso.club-internet.fr/plafaye`

# DOCUMENTS SPÉCIAUX

Une disquette contenant les programmes informatiques est annexée à ce document.

**RÉSUMÉ :** On construit un test d'ajustement de la normalité pour les innovations d'un modèle ARMA($p$, $q$) de tendance et moyenne connues, basé sur l'approche du test lisse dépendant des données et simple à appliquer. Une vaste étude de simulation est menée pour étudier ce test pour des tailles échantillonnales modérées. Notre approche est en général plus puissante que les tests existants. Le niveau est tenu sur la majeure partie de l'espace paramétrique. Cela est en accord avec les résultats théoriques montrant la supériorité de l'approche du test lisse dépendant des données dans des contextes similaires.
Un test d'indépendance (ou d'indépendance sérielle) semi-paramétrique entre des sous-vecteurs de loi normale est proposé, mais sans supposer la normalité jointe de ces marginales. La statistique de test est une fonctionnelle de type Cramér-von Mises d'un processus défini à partir de la fonction caractéristique empirique. Ce processus est défini de façon similaire à celui de Ghoudi et al. (2001) construit à partir de la fonction de répartition empirique et utilisé pour tester l'indépendance entre des marginales univariées. La statistique de test peut être représentée comme une V-statistique. Il est convergent pour détecter toute forme de dépendance. La convergence faible du processus est établie. La distribution asymptotique des fonctionnelles de Cramér-von Mises est approchée par la méthode de Cornish-Fisher au moyen d'une formule de récurrence pour les cumulants et par le calcul numérique des valeurs propres dans la formule d'inversion. La statistique de test est comparée avec celle de Wilks pour l'hypothèse paramétrique d'indépendance dans le modèle MANOVA à un facteur avec effets aléatoires.

**MOTS-CLÉS :** Processus ARMA, Bruit blanc Gaussien, Test d'ajustement, Normalité des résidus, Test lisse, Fonction caractéristique, Indépendance, Analyse multivariée, Indépendance sérielle, Processus stochastiques

**TITLE :** Independence tests in multivariate analysis and normality tests in ARMA models.

**ABSTRACT :** We build a goodness-of-fit test of normality for the innovations of an ARMA($p$, $q$) model with known mean or trend. This test is based on the data driven smooth test approach and is simple to perform. An extensive simulation study is conducted to study the behavior of the test for moderate sample sizes. Our approach is generally more powerful than existing tests while holding its level throughout most of the parameter space. This agrees with theoretical results showing the superiority of the data driven smooth test approach in related contexts.
A semi-parametric test of independence (or serial independence) is proposed between marginal vectors each of which is normally distributed but without assuming the joint normality of these marginal vectors. The test statistic is a Cramér-von Mises functional of a process defined from the empirical characteristic function. This process is defined similarly as the process of Ghoudi et al. (2001) built from the empirical distribution function and used to test for independence between univariate marginal variables. The test statistic can be represented as a V statistic. It is consistent to detect any form of dependence. The weak convergence of the process is derived. The asymptotic distribution of the Cramér-von Mises functionals is approximated by the Cornish-Fisher expansion using a recursive formula for cumulants and by the numerical evaluations of the eigenvalues in the inversion formula. The test statistic is finally compared with Wilks' statistic for testing the parametric hypothesis of independence in the one-way MANOVA model with random effects.

**KEYWORDS :** ARMA process, Gaussian white noise, Normality of residuals, Smooth test, Characteristic function, Independence, Multivariate Analysis, Serial independence, Stochastic processes

**DISCIPLINE :** Mathématiques Appliquées : Statistique

**LABORATOIRES :** Laboratoire de probabilités et statistique - Département des sciences mathématiques - Université Montpellier II - Place Eugène Bataillon, 34095 Montpellier Cedex 5, France ET Département de mathématiques et de statistique - Université de Montréal - CP 6128 succ. Centre-Ville, Montréal QC H3C 3J7, Canada