Université de Montréal

Estimation et validation de modèles non-linéaires multivariés dans l'analyse des séries chronologiques

par

Dominique Chabot-Hallé

Département de mathématiques et de statistique
Faculté des arts et des sciences

Mémoire présenté à la Faculté des études supérieures
en vue de l'obtention du grade de
Maître ès sciences (M.Sc.)
en statistique

août 2007

Université de Montréal

Faculté des études supérieures

Ce mémoire intitulé

Estimation et validation de modèles non-linéaires multivariés dans l'analyse des séries chronologiques

présenté par

Dominique Chabot-Hallé

a été évalué par un jury composé des personnes suivantes :

Yves Lepage

(président-rapporteur)

Pierre Duchesne

(directeur de recherche)

Louis Doray

(membre du jury)

Mémoire accepté le:

27 juin 2007

SOMMAIRE

Le présent mémoire traite d'une classe très générale de modèles non-linéaires multivariés. Dans un premier temps, nous discutons d'une méthode d'estimation dans cette classe de modèles. Ensuite, nous obtenons la distribution asymptotique des autocovariances et autocorrélations résiduelles dans cette classe considérant que le terme d'erreur est une différence de martingale. Nous proposons ensuite deux tests diagnostiques : un test de type portemanteau et un test pour les délais individuels. Pour illustrer les contributions proposées dans ce mémoire, une étude de simulations est effectuée où nous estimons et validons un modèle non-linéaire multivarié. Dans cette étude nous comparons, en considérant plusieurs termes d'erreur, la statistique de test supposant l'indépendance du terme d'erreur à celle plus générale supposant uniquement que le terme d'erreur est une différence de martingale.

Mots-clés : modèles non-linéaires multivariés, différence de martingale, tests diagnostiques, tests de type portemanteau.

SUMMARY

In this master's thesis we discuss a very general class of multivariate nonlinear models for time series analysis. First we discuss an estimation method for this class of models. Then, we derive the asymptotic distribution of residual autocovariance and autocorrelation matrices for this class assuming only that the error term is a martingale difference sequence. Two types of applications are developed : global test statistics of the portmanteau type and one-lag test statistics. To illustrate the proposed methodology, a simulation study was conducted by estimating and diagnosing a non-linear multivariate model. In this study we compare, considering different error terms, the classical test statistics assuming independent errors and the proposed methodology which supposes only martingale difference errors.

Key words : nonlinear multivariate models, autocovariance matrices, diagnostic checking, martingale difference sequence, portmanteau test statistics.

TABLE DES MATIÈRES

LISTE DES FIGURES

LISTE DES TABLEAUX

REMERCIEMENTS

Pour ma fille.

Dans un premier temps, je tiens à remercier mon directeur de recherche, Pierre Duchesne, pour son aide précieuse, sa grande disponibilité, sa patience et non le moindre attribut, son sens de l'humour. Je tiens également à le remercier pour son aide financière. Ensuite, je tiens à remercier ma famille : mon épouse Swan pour ses encouragements son aide et sa patience, ma mère pour l'amour inconditionnel quelle me porte, mon père pour ses conseils judicieux et bien sûr, le reste ma famille, merci. Je tiens également à remercier le personnel administratif et enseignant de l'Université de Montréal qui ont grandement facilité mon travail ainsi que mes confrères et amis qui ont rendu cette expérience tellement enrichissante.

INTRODUCTION

Depuis le début des années soixante, l'étude des séries chronologiques a connu un grand bond avec l'arrivé des modèles autorégressifs-moyennes-mobiles (ARMA) de Box et Jenkins (voir l'ouvrage de Box, Jenkins et Reinsel (1994)). Les modèles ARMA comportent beaucoup d'avantages. Ils sont relativement simples à analyser d'un point de vue mathématique et une théorie complète de ces modèles est déjà élaborée. De plus, ils sont faciles à estimer et des outils informatiques sont disponibles pour les valider adéquatement. Bien entendu, les modèles ARMA ont également certains points faibles. Tong (1990) en mentionne plusieurs. En particulier, il est reconnu que les modèles ARMA ne conviennent pas pour modéliser des données asymétriques ou comportant des sauts importants. De plus, les modèles ARMA semblent mal adaptés pour des données fortement cycliques. Ces désavantages des modèles ARMA peuvent être en partie expliqués par une de leurs caractéristiques : les propriétés de ces modèles sont intimement reliées aux propriétés des modèles linéaires qui reposent souvent sur l'écriture sous la forme d'un processus linéaire avec erreurs indépendantes. Or, dans de nombreux domaines de recherche, les phénomèmes mesurés dans le temps admettent une forme non-linéaire. Entre autres, il est admis que la plupart des variables économiques affichent un comportement non-linéaire (voir Granger et Teräsvirta (1993)). Souvent, ces modèles non-linéaires peuvent s'exprimer suivant une écriture linéaire par le Théorème de Représentation de Wold, s'ils satisfont la propriété de stationnarité. Cependant, en général, le terme d'innovation n'est alors plus indépendant ; il ne satisfait que la propriété d'absence de corrélation. Par conséquent, les modèles non-linéaires connaissent aujourd'hui une popularité sans précédent vu l'impossibilité de modéliser adéquatement certaines séries chronologiques à l'aide des modèles ARMA.

Soit $\mathbf{Y} = \{\mathbf{Y}_t, t \in \mathbb{Z}\}$ un processus stochastique multivarié strictement stationnaire et ergodique. Les définitions précises de stationnarité et d'ergodicité sont discutées au premier chapitre. Considérons la classe des modèles non-linéaires multivariés suivante :

$$\mathbf{Y}_t = \mathbf{f}(\mathcal{I}_{t-1}; \boldsymbol{\theta}_0) + \mathbf{a}_t, \tag{0.1}$$

où $\mathcal{I}_t$ représente $\sigma(\mathbf{Y}_t, \mathbf{Y}_{t-1}, \ldots)$, la sigma-algèbre générée par $\{\mathbf{Y}_t, \mathbf{Y}_{t-1}, \ldots\}$, $\mathbf{f} = (f_1, \ldots, f_d)^\top$ est une fonction connue à valeurs réelles dans $\mathbb{R}^d$, et $\boldsymbol{\theta}_0$ est un vecteur de dimension $K \times 1$ des paramètres inconnus. Supposons que $f_i \equiv f_i(\mathcal{I}_{t-1}; \boldsymbol{\theta})$, $i = 1, \ldots, d$, est une fonction des $\mathbf{Y}_t$ passés et de $\boldsymbol{\theta}$. De plus, supposons que la fonction vectorielle $\mathbf{f}(\cdot;\cdot)$ admette une dérivée seconde continue par rapport à $\boldsymbol{\theta}$, presque partout. Le terme d'erreur $\mathbf{a} = \{\mathbf{a}_t, t \in \mathbb{Z}\}$ est une différence de martingale, où $\mathbf{a}_t = (a_t(1), \ldots, a_t(d))^\top$, $t \in \mathbb{Z}$. On note que $\mathbf{a}$ forme une suite de vecteurs aléatoires de moyennes nulles non-corrélés, mais pas nécessairement indépendants. Un très grand nombre de modèles admettent la forme de l'équation (0.1). Entre autres, on retrouve dans les modèles univariés les modèles bilinéaires proposés initialement par Granger et Anderson (1978), les modèles autorégressifs conditionnellement hétéroscédastiques (ARCH) de Engle (1982), et leurs généralisations (par exemple les ARCH généralisés ou GARCH, étudiés par Bollerslev (1986)), les modèles autorégressifs à seuils étudiés par Tong (1978, 1983), notés TAR, et les versions de type SETAR (pour *self-exciting threshold autoregressive*) et les modèles autorégressifs avec fonctions de transition lisse (STAR) étudiés en détails dans Teräsvirta (1994). Dans un cadre multivarié, on retrouve la version multivarié des ARMA, souvent dénotées VARMA. Naturellement, les VARMA possèdent les mêmes limites que les ARMA : ils peuvent modéliser adéquatement des phénomènes intrinsèquement linéaires. Une référence générale est Brockwell et Davis (2002). Des extensions des modèles à seuils et des modèles STAR pour séries chronologiques multivariés se retrouvent dans Tsay (1998) et van Dijk, Teräsvirta et Franses (2002).

Pour pouvoir modéliser une série chronologique comme dans l'équation (0.1), il faut avant tout disposer des outils nécessaires afin d'estimer le modèle. Dans un premier temps, nous présentons la théorie asymptotique disponible sur les estimateurs des paramètres de (0.1). Klimko et Nelson (1978) ont étudié le comportement des estimateurs des moindres carrés conditionnels dans le cadre des modèles non-linéaires univariés. Tjøstheim (1986) a généralisé ce résultat au cas multivarié en permettant l'utilisation d'un critère d'estimation de forme générale. Ensuite, nous nous intéressons à un outil de base dans la validation des modèles en séries chronologiques : les autocovariances résiduelles. Li (2004) présente un survol des différentes méthodes de validation en séries chronologiques. De nombreux résultats ont été obtenus pour des classes particulières de modèles non-linéaires. Par exemple, Duchesne (2004) obtient la distribution asymptotique d'un certain type d'autocovariances résiduelles pour la classe des modèles ARCH vectoriels, permettant ainsi de développer des tests de validation de structure ARCH pour cette classe. Cependant, il est possible de regrouper un grand nombre de modèles non-linéaires dans une classe plus générale. Li (1992) a obtenu la distribution asymptotique des autocovariances résiduelles pour une classe très générale de modèles non-linéaires univariés avec terme d'erreur indépendant. Les accomplissements principaux présentés dans ce mémoire sont la généralisation du résultat de Li (1992) au cas multivarié ainsi que l'assouplissement de l'hypothèse d'indépendance du terme d'erreur, en supposant uniquement que le terme d'erreur est une différence de martingale. À partir de ce résultat, deux types de statistiques de tests sont développés : des tests de type portemanteau et des tests pour les délais individuels.

Afin d'illustrer nos résultats théoriques, nous avons effectué une étude de simulations où nous nous intéressons aux niveaux empiriques des deux statistiques en considérant un certain modèle non-linéaire multivarié. Nous avons également considéré différents termes d'erreurs pour mettre en évidence la différence entre la statistique de test supposant l'indépendance du terme d'erreur à celle, plus générale, supposant uniquement que le terme d'erreur est une différence de martingale. Le présent mémoire est organisé comme suit. Dans le premier chapitre, des préliminaires sont donnés sur les modèles non-linéaires multivariés. Le deuxième chapitre contient une copie intégrale d'un article actuellement soumis pour publication qui développe nos statistiques de type portemanteau. En annexe, des résultats complémentaires de simulation sont présentés.

Chapitre 1

PRÉLIMINAIRES

1.1. Séries chronologiques multivariées

Dans cette première section, nous allons nous concentrer sur les concepts de base en séries chronologiques multivariées. Une série chronologique multivariée de taille n est définie comme une réalisation finie d'un processus stochastique multivarié $\mathbf{Y} = \{\mathbf{Y}_t, t \in \mathbb{Z}\}$, où le vecteur aléatoire $\mathbf{Y}_t = (Y_t(1), \ldots, Y_t(d))^\top$ est de dimension d. Pour une telle série multivariée, présumant $E\{Y_t^2(i)\} < \infty$ pour tout t et tout i, on peut définir ses deux premiers moments. On définit le vecteur des moyennes :

$$E(\mathbf{Y}_t) = \boldsymbol{\mu}_t = (\mu_t(1), \ldots, \mu_t(d))^\top, \tag{1.1.1}$$

ainsi que les matrices de covariance :

$$E\{(\mathbf{Y}_t - \boldsymbol{\mu}_t)(\mathbf{Y}_{t-h} - \boldsymbol{\mu}_{t-h})^\top\} = \boldsymbol{\Gamma}(t, t-h) = [\gamma_{ij}(t, t-h)]_{i,j=1}^d. \tag{1.1.2}$$

Les covariances croisées $\gamma_{ij}(t, t-h)$ donnent une indication de la dépendance entre les observations de différentes séries (lorsque $i \neq j$).

1.2. Processus stationnaires et ergodicité

Lors de l'analyse des séries chronologiques, le concept de processus stationnaire joue un rôle fort important. On commence ainsi par définir la stationnarité au sens strict.

Définition 1.2.1. *Un processus $\{\mathbf{Y}_t, t \in \mathbb{Z}\}$ de dimension d est dit strictement stationnaire si, pour tout $n \in \mathbb{N}$, $t_1, \cdots, t_n$, $h \in \mathbb{Z}$, la distribution de $\mathbf{Y}_{t_1}, \cdots, \mathbf{Y}_{t_n}$ et celle de $\mathbf{Y}_{t_1+h}, \cdots, \mathbf{Y}_{t_n+h}$ sont les mêmes.*

Pour un processus stochastique donné, vérifier la condition de stationnarité au sens strict est souvent difficile. C'est particulièrement vrai pour les processus non-linéaires. Lasota et Mackey (1989) ont simplifié cette tâche en démontrant le résultat suivant.

Théorème 1.2.1. *Soit $\{\mathbf{Y}_t, t \in \mathbb{Z}\}$ une chaine de Markov de dimension d définie par :*

$$\mathbf{Y}_t = \mathbf{f}(\mathbf{Y}_{t-1}, a_t), t = 1, 2, \ldots$$

De plus, supposons que $\mathbf{Y}_t$ vérifie les hypothèses $(A1) - (A4)$:

(A1) $\mathbf{f}(\mathbf{y}, a)$ *est continue sur* $\mathbf{y} \in \mathbb{R}^d$.

(A2) Le processus $\{a_t, t \geq 1\}$ est une suite de variables aléatoires indépendantes et identiquement distribués (iid). Le processus $\{a_t\}$ est supposé indépendant de $\mathbf{Y}_0$.

(A3) $E\{\| \mathbf{f}(\mathbf{y}, a) - \mathbf{f}(\mathbf{x}, a) \|\} < \| \mathbf{y} - \mathbf{x} \|$ *pour* $\mathbf{y}, \mathbf{x} \in \mathbb{R}^d$, $\mathbf{y} \neq \mathbf{x}$, *où* $\| \cdot \|$ *représente la norme euclidienne.*

(A4) $E\{\| \mathbf{f}(\mathbf{y}, a) \|^2\} \leq \alpha \| \mathbf{y} \|^2 + \beta$ *pour* $\mathbf{y} \in \mathbb{R}^d$, *où α et β sont des constantes nonnégatives et $\alpha < 1$.*

Alors, $\{\mathbf{Y}_t\}$ est strictement stationnaire.

À l'aide de ce résultat, il devient beaucoup plus facile de trouver, pour un modèle donné, les conditions sur les paramètres afin de s'assurer de la stationnarité au sens strict. Par exemple, Taniguchi et Kakizawa (2000, pp. 86-87) utilisent ce théorème pour trouver l'espace des paramètres impliquant la stationnarité au sens strict pour quelques modèles non-linéaires.

Une condition moins restrictive que la stationnarité au sens strict et plus facilement vérifiable en pratique est la stationnarité au sens large.

Définition 1.2.2. *Un processus $\{\mathbf{Y}_t, t \in \mathbb{Z}\}$ de dimension d, tel que $E\{Y_t^2(i)\} < \infty$ pour tout t et tout i, est stationnaire au sens large si le premier et les deuxièmes moments sont invariants dans le temps. Le premier et les seconds moments définis* *respectivement dans l'équation (1.1.1) et (1.1.2) se réduisent à :*

$$E(\mathbf{Y}_t) = \boldsymbol{\mu},\ \forall t,$$

$$E\left\{(\mathbf{Y}_t - \boldsymbol{\mu})(\mathbf{Y}_{t-h} - \boldsymbol{\mu})^\top\right\} = \boldsymbol{\Gamma}(h),\ \forall h \in \mathbb{Z}.$$

Une autre propriété importante dans l'étude des processus stochastiques est l'ergodicité. La définition de l'ergodicité est quelque peu technique et nécessite un cadre probabiliste bien défini. Voir Taniguchi et Kakizawa (2000, p. 18), qui en discutent plus longuement. Il est cependant important de noter que si un processus stochastique est strictement stationnaire et ergodique, alors un ensemble de théorèmes devient disponible pour l'étude des propriétés asymptotiques de plusieurs quantités d'intérêt. Ces théorèmes, souvent appelés théorèmes d'ergodicité, sont en quelque sorte des théorèmes de type "lois des grands nombres" dans le contexte des processus stochastiques. Chan et Tong (1986) discutent des conditions de stationnarité et d'ergodicité pour la classe des processus admettant une structure SETAR. Les processus SETAR sont décrits plus longuement à la Section 1.3.2.

Le terme d'erreur $\{\mathbf{a}_t\}$ du modèle général décrit dans l'équation (0.1) mérite d'être discuté en détails. De manière générale, le terme d'erreur $\{\mathbf{a}_t\}$ pourrait être un bruit blanc faible. Ceci fait l'objet de la Définition 1.2.3.

Définition 1.2.3. *Soit $\{\mathbf{a}_t, t \in \mathbb{Z}\}$ un processus stochastique qui satisfait les conditions suivantes :*

(i) $E(\mathbf{a}_t) = \mathbf{0}$,

(ii) $E(\mathbf{a}_t\mathbf{a}_t^\top) = \boldsymbol{\Sigma}_{\mathbf{a}}$,

(iii) $E(\mathbf{a}_t\mathbf{a}_s^\top) = \mathbf{0},\ \forall t \neq s$,

alors $\{\mathbf{a}_t\}$ est un bruit blanc faible.

On note dans la définition précédente que $\{\mathbf{a}_t\}$ n'est pas forcément constitué de vecteurs indépendants observés dans le temps. Lorsque le processus $\{\mathbf{a}_t\}$ est constitué de vecteurs aléatoires iid de moyennes nulles et de variance commune $\boldsymbol{\Sigma}_{\mathbf{a}}$, on dit alors que $\{\mathbf{a}_t\}$ est un bruit blanc fort.

L'hypothèse de bruit blanc faible est très générale et celle de bruit blanc fort trop restrictive. Nous allons présumer que le terme d'erreur $\{\mathbf{a}_t\}$ du modèle (0.1) est une différence de martingale. Nous définissons plus précisément ce concept dans la définition suivante :

Définition 1.2.4. *Soit* $\{\mathbf{Y}_t, t \in \mathbb{Z}\}$ *un processus stochastique qui satisfait :*

$$E(\mathbf{Y}_t|\mathcal{I}_{t-1}) = \mathbf{0} \text{ presque partout pour tout } t \in \mathbb{N},$$

alors $\{\mathbf{Y}_t\}$ *est une différence de martingale.*

Lorsque $\{\mathbf{a}_t\}$ est une différence de martingale, il est facile de vérifier qu'en fait les vecteurs aléatoires $\mathbf{a}_t$ et $\mathbf{a}_s$ sont non-corrélés. Il est à noter cependant qu'il est possible qu'un bruit blanc faible ne satisfasse pas l'hypothèse de différence de martingale. Cependant, ce qui rend particulièrement intéressant un terme d'erreur satisfaisant la Définition 1.2.4, est qu'il est possible que $\mathbf{a}_t$ et $\mathbf{a}_s$ soient non-corrélés mais dépendants. Par exemple, si on pose $\mathbf{a}_t = \mathbf{b}_t b_{t-1}(1)$, $\mathbf{b}_t = (b_t(1), \ldots, b_t(d))^\top$, $\{\mathbf{b}_t\}$ étant un bruit blanc fort, alors $\{\mathbf{a}_t\}$ est une différence de martingale.

Nous énonçons sans démonstration deux théorèmes fortement utilisés. Le premier est en fait un théorème de type "lois des grands nombres" lorsque $\{\mathbf{Y}_t\}$ est strictement stationnaire et ergodique. Le second est un théorème central limite lorsque $\{\mathbf{Y}_t\}$ est une différence de martingale.

Théorème 1.2.2. *Soit* $\{\mathbf{Y}_t, t \in \mathbb{Z}\}$ *un processus strictement stationnaire et ergodique tel que* $E||\mathbf{Y}_t|| < \infty$*, alors :*

$$\frac{1}{n}\sum_{t=1}^{n} \mathbf{Y}_t \xrightarrow{p.s} E(\mathbf{Y}_1).$$

De plus, si $E||\mathbf{Y}_t||^2 < \infty$*, alors :*

$$\frac{1}{n}\sum_{t=1}^{n} \mathbf{Y}_t\mathbf{Y}_{t+m}^\top \xrightarrow{p.s} E(\mathbf{Y}_1\mathbf{Y}_{1+m}^\top).$$

La preuve est donnée dans Taniguchi et Kakizawa (2000, p. 19). Le théorème suivant est dû à Ibragimov (1963).

Théorème 1.2.3. *Soit* $\{Y_t, t \in \mathbb{Z}\}$ *un processus strictement stationnaire et ergodique tel que* $E(Y_t^2) = \sigma^2, E(Y_t|\mathcal{I}_{t-1}) = 0$*, p.s. Alors,*

$$\frac{1}{\sigma\sqrt{n}}\sum_{t=1}^{n} Y_t \xrightarrow{d} N(0,1).$$

Ces théorèmes sont présentés en détails dans Stout (1974).

1.3. Modèle non-linéaire multivarié

Nous étudions plus en détails quelques modèles non-linéaires, respectant l'écriture (0.1), qui sont souvent considérés dans la littérature.

1.3.1. Quelques exemples de modèles

Une classe de modèles fort populaires est celle des modèles à changement de régimes. Ces modèles ont la particularité de changer d'états dans le temps. Les différents états de la série chronologique sont appelés régimes. Tong (1983) donne de nombreux exemples provenant de domaines divers où le concept de régime joue un rôle important. Parmis ces derniers, il mentionne l'océanographie, l'hydrologie, l'économie et le génie médical. Teräsvirta (1994) a popularisé un de ces modèles en particulier. Un processus STAR(p) vectoriel à d dimensions est défini comme suit :

Définition 1.3.1.

$$\begin{aligned}\mathbf{Y}_t &= (\mathbf{\Phi}_{1,1} + \mathbf{\Phi}_{1,2}\mathbf{Y}_{t-1} + \ldots + \mathbf{\Phi}_{1,p}\mathbf{Y}_{t-p})\{1 - G(s_t;\gamma;c)\} \\ &\quad + (\mathbf{\Phi}_{2,1} + \mathbf{\Phi}_{2,2}\mathbf{Y}_{t-1} + \ldots + \mathbf{\Phi}_{2,p}\mathbf{Y}_{t-p})G(s_t;\gamma;c) + \mathbf{a}_t,\end{aligned}$$

où $\mathbf{\Phi}_{1,1}, \ldots, \mathbf{\Phi}_{1,p}$ *sont les matrices* $d \times d$ *de coefficients du premier régime,* $\mathbf{\Phi}_{2,1}, \ldots, \mathbf{\Phi}_{2,p}$ *sont les matrices* $d \times d$ *de coefficients du second régime,* G *est la fonction de transition et* s_t *est la variable aléatoire de transition.*

Pour la fonction de transition G, Teräsvirta (1994) préconise deux choix. Soit le modèle LSTAR (STAR logistique), si on considère la fonction de transition suivante :

$$G(s_t;\gamma;c) = [1 + \exp\{-\gamma(s_t - c)\}]^{-1}, \gamma > 0,$$

où γ est proportionelle à la vitesse de la transition et c permet de centrer la transition, ou encore le modèle ESTAR (STAR exponentielle), si on prend la fonction de transition qui suit :

$$G(s_t;\gamma;c) = 1 - \exp\{-\gamma(s_t - c)\}^2, \gamma > 0.$$

Van Dijk, Teräsvirta et Franses (2002) discutent des différences entre ces deux fonctions de transition. On remarque que le modèle présenté dans la Définition 1.3.1 considère une fonction de transition G unique aux d séries. Il est possible de généraliser ce modèle pour permettre une fonction de transition G pour chaque série. On remarque également que la variable de transition s_t, qui est généralement une fonction linéaire des valeurs retardées du processus $\{\mathbf{Y}_t\}$, est la même pour les d séries. Il est également possible de généraliser ce modèle pour permettre l'utilisation de différentes variables de transition pour les séries. van Dijk, Teräsvirta et Franses (2002) discutent de ces différentes possibilités. Cependant, on note que la modélisation de données à l'aide de modèles STAR vectoriels semble intuitivement appropriée pour des séries chronologiques où la nature de la transition est la même. Pour citer un exemple, on peut penser à deux séries économiques avec un état de récession et un état d'expansion économique. De plus, considérer plusieurs fonctions de transition ainsi que plusieurs variables de transition nécessite un très grand nombre de paramètres à estimer qui risquent de causer une surparamétrisation.

Le modèle SETAR est un cas particulier du modèle STAR où la fonction de transition est la fonction indicatrice. Ce modèle a été étudié par Tong (1978, 1983). Tsay (1998) fournit un exemple de modélisation de données réelles avec un modèle SETAR multivarié. Les processus SETAR(p) vectoriels à d dimensions sont définis dans la Définition 1.3.2 qui suit.

Définition 1.3.2.

$$\mathbf{Y}_t = (\mathbf{\Phi}_{1,1}\mathbf{Y}_{t-1}+\ldots+\mathbf{\Phi}_{1,p}\mathbf{Y}_{t-p})\{1-I(s_t)\}+(\mathbf{\Phi}_{2,1}\mathbf{Y}_{t-1}+\ldots+\mathbf{\Phi}_{2,p}\mathbf{Y}_{t-p})I(s_t)+\mathbf{a}_t,$$

où

$$I(x) = \begin{cases} 1, & si\ x > 0, \\ 0, & sinon, \end{cases}$$

$\mathbf{\Phi}_{1,1},\ldots,\mathbf{\Phi}_{1,p}$ *sont les matrices* $d\times d$ *de coefficients du premier régime,* $\mathbf{\Phi}_{2,1},\ldots,\mathbf{\Phi}_{2,p}$ *sont les matrices* $d\times d$ *de coefficients du second régime et* s_t *est la variable aléatoire de transition.*

Une tout autre classe de modèles fort populaire est celle des modèles ARCH. Engle (1982) a introduit cette classe de modèles dans le but de mieux expliquer certains phénomènes économiques. Cette contribution lui a valu le prix Nobel de 2003 en économie. Depuis son introduction, la classe des modèles ARCH s'est démarquée par son aptitude à modéliser adéquatement des données financières. On dit qu'un processus univarié est ARCH(p) s'il admet la représentation donnée dans la définition qui suit :

Définition 1.3.3.

$$\begin{aligned} Y_t &= u_t h_t^{1/2}, \\ h_t &= b_0 + \sum_{i=1}^{p} b_i Y_{t-i}^2. \end{aligned}$$

où $b_0 > 0$, $b_i \geq 0$, $i = 1, \ldots, p$, *et* $\{u_t\}$ *est un bruit blanc fort univarié.*

On remarque que dans ce modèle $\{Y_t\}$ est dépendant mais non-corrélé. On note aussi que $\{Y_t\}$ est une différence de martingale. Pour des versions multivariées des modèles ARCH(p), ou de manière plus générale des modèles GARCH(p,q), voir Engle et Kroner (1995).

1.4. Estimation des modèles non-linéaires multivariés

Estimer les paramètres d'un modèle non-linéaire nécessite la même approche que pour un modèle linéaire, c'est-à-dire qu'on s'intéresse à un critère que l'on cherche à minimiser par rapport aux paramètres. Si notre critère est judicieusement choisi, alors les estimateurs de nos paramètres auront des propriétés souhaitables. Par exemple, dans le cadre de la régression linéaire, sous l'hypothèse que les erreurs sont gaussiennes, minimiser le critère des moindres carrés donne des estimateurs sans biais à variance minimale. De plus, dans ce contexte, il est possible d'obtenir une expression analytique pour les estimateurs. Pour un modèle non-linéaire, obtenir une expression analytique pour les estimateurs se révèle à toutes fins utiles impensable. La solution habituelle consiste à minimiser un critère d'intérêt et la recherche des estimateurs s'effectue en pratique à l'aide d'algorithmes d'optimisation. Dans le cadre de ce projet, nous avons préconisé l'utilisation de la technique des moindres carrés et avons trouvé les estimateurs en considérant l'algorithme de Gauss-Newton.

Pour pouvoir effectuer de l'inférence sur nos paramètres, il est également important d'obtenir la distribution asymptotique des estimateurs des paramètres. Considérons le modèle non-linéaire multivarié de dimension d défini dans l'équation (0.1). Définissons le critère des moindres carrés comme suit :

$$S_n(\boldsymbol{\theta}) = \sum_{t=\tau+1}^{n} \{\mathbf{Y}_t - \mathbf{f}(\mathcal{I}_{t-1}; \boldsymbol{\theta})\}^\top \boldsymbol{\Sigma}_{\mathbf{a}}^{-1} \{\mathbf{Y}_t - \mathbf{f}(\mathcal{I}_{t-1}; \boldsymbol{\theta})\},$$

où τ est une constante adéquatement choisie. Voir Taniguchi et Kakizawa (2000, p. 98) pour une discussion du choix de cette constante. Il est possible d'obtenir la distribution asymptotique des estimateurs $\hat{\boldsymbol{\theta}}_n$ de $\boldsymbol{\theta}$ qui minimisent $S_n(\boldsymbol{\theta})$. La preuve complète du résultat suivant est fournit dans l'annexe A.1. La démonstration s'inspire de Tjøstheim (1986).

Théorème 1.4.1. *Soit $\{\mathbf{Y}_t, t \in \mathbb{Z}\}$ un processus strictement stationnaire et ergodique tel que $E\{||\mathbf{Y}_t||^2\} < \infty$ et où $\mathbf{f}(\mathcal{I}_{t-1}, \boldsymbol{\theta})$ est trois fois différentiable p.s. sur un ensemble ouvert B contenant la vraie valeur $\boldsymbol{\theta}_0$. De plus, on suppose que les conditions suivantes sont respectées :*

(C1) Pour $j, l = 1, \ldots, K$:

$$E\left[\left|\left|\frac{\partial}{\partial \theta_j}\left\{\mathbf{f}(\mathcal{I}_{t-1}, \boldsymbol{\theta}_0)^\top \boldsymbol{\Sigma}_{\mathbf{a}}^{-1} \mathbf{f}(\mathcal{I}_{t-1}, \boldsymbol{\theta}_0)\right\}\right|\right|\right] < \infty,$$

et

$$E\left[\left|\left|\frac{\partial^2}{\partial \theta_j \partial \theta_l}\left\{\mathbf{f}(\mathcal{I}_{t-1}, \boldsymbol{\theta}_0)^\top \boldsymbol{\Sigma}_{\mathbf{a}}^{-1} \mathbf{f}(\mathcal{I}_{t-1}, \boldsymbol{\theta}_0)\right\}\right|\right|\right] < \infty.$$

(C2) Pour $j = 1, \ldots, K$, les vecteurs

$$\boldsymbol{\Sigma}_{\mathbf{a}}^{-1/2} \frac{\partial \mathbf{f}(\mathcal{I}_{t-1}, \boldsymbol{\theta}_0)}{\partial \theta_j},$$

sont linéairement indépendants au sens de l'algèbre linéaire.

(C3) Pour $\boldsymbol{\theta} \in B$, il existe une fonction $H_t^{ijl}(\mathbf{Y}_1, \ldots., \mathbf{Y}_{t-1})$ telle que :

$$\left|\frac{\partial^3}{\partial \theta_i \partial \theta_j \partial \theta_l}\left\{\mathbf{f}(\mathcal{I}_{t-1}, \boldsymbol{\theta})^\top \boldsymbol{\Sigma}_{\mathbf{a}}^{-1} \mathbf{f}(\mathcal{I}_{t-1}, \boldsymbol{\theta})\right\}\right| \leq H_t^{ijl},$$

avec

$$E(H_t^{ijl}) < \infty,$$

pour $i, j, l = 1, \ldots, K$.

(C4) De plus, la matrice suivante existe :

$$\mathbf{R} = E\left[\left\{\frac{\partial \mathbf{f}(\mathcal{I}_{t-1}, \boldsymbol{\theta}_0)^\top}{\partial \boldsymbol{\theta}} \boldsymbol{\Sigma}_{\mathbf{a}}^{-1} \mathbf{a}_t\right\} \left\{\mathbf{a}_t^\top \boldsymbol{\Sigma}_{\mathbf{a}}^{-1} \frac{\partial \mathbf{f}(\mathcal{I}_{t-1}, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}^\top}\right\}\right] < \infty.$$

Alors il existe une suite d'estimateurs $\hat{\boldsymbol{\theta}}_n$ *telle que* $\hat{\boldsymbol{\theta}}_n \xrightarrow{p.s} \boldsymbol{\theta}_0$, *et pour tout* $\varepsilon > 0$, *il existe un évènement* E *avec* $P(E) > 1 - \varepsilon$ *et un* n_0 *tel que sur* E, *pour* $n > n_0$, $\partial S_n(\hat{\boldsymbol{\theta}}_n)/\partial \boldsymbol{\theta} = 0$ *et* $S_n(\boldsymbol{\theta})$ *atteint un minimum relatif à* $\hat{\boldsymbol{\theta}}_n$. *De plus,*

$$n^{1/2}\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right) \xrightarrow{d} N(\mathbf{0}, \mathbf{U}^{-1}\mathbf{R}\mathbf{U}^{-1}),$$

où

$$\mathbf{U} = E\left\{\frac{\partial \mathbf{f}(\mathcal{I}_{t-1}, \boldsymbol{\theta}_0)^\top}{\partial \boldsymbol{\theta}} \boldsymbol{\Sigma}_{\mathbf{a}}^{-1} \frac{\partial \mathbf{f}(\mathcal{I}_{t-1}, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}^\top}\right\},$$

et $\mathbf{R}$ *est défini plus haut.*

Avec ce résultat, on peut estimer la matrice de covariance des estimateurs, en estimant les espérances mathématiques par des moyennes échantillonnales et en remplaçant les paramètres inconnus par les estimateurs des moindres carrés. On peut donc construire des intervalles de confiance pour les paramètres ou encore effectuer des tests d'hypothèses. Voir Gallant (1987) pour plus de détails.

1.5. Diagnostics des séries chronologiques multivariées

1.5.1. Estimation des autocovariances matricielles

De manière générale, lorsque l'on étudie une série chronologique, on ne dispose pas de la vraie fonction d'autocovariance. Il est cependant possible de développer une version échantillonnale de la fonction d'autocovariance théorique. Posons :

$$\hat{\boldsymbol{\Gamma}}_{\mathbf{Y}}(l) = \begin{cases} n^{-1}\sum_{t=l+1}^{n}(\mathbf{Y}_t - \bar{\mathbf{Y}})(\mathbf{Y}_{t-l} - \bar{\mathbf{Y}})^\top, & l = 0, 1, \ldots, n-1, \\ \hat{\boldsymbol{\Gamma}}_{\mathbf{Y}}^\top(-l), & l = -1, -2, \ldots, -n+1, \end{cases}$$

où $\bar{\mathbf{Y}} = n^{-1}\sum_{t=1}^{n}\mathbf{Y}_t$. La fonction $\{\hat{\boldsymbol{\Gamma}}_{\mathbf{Y}}(l), |l| < n\}$ est appelée la fonction d'autocovariance échantillonnale. Les propriétés générales des autocovariances échantillonnales sont décrites dans Hannan (1970) lorsque $\{\mathbf{Y}_t\}$ est un processus stochastique stationnaire multivarié.

1.5.2. Les autocovariances résiduelles matricielles

Un outil de base en séries chronologiques pour diagnostiquer un modèle repose sur l'utilisation des autocovariances résiduelles. Dans un tel cas, pour un terme d'erreur $\{\mathbf{a}_t\}$, l'hypothèse d'adéquation revient souvent à tester que $\mathbf{\Gamma}_{\mathbf{a}}(l) = \mathbf{0}$, $l = 1, 2, \ldots$, où $\mathbf{\Gamma}_{\mathbf{a}}(l)$ dénote la matrice d'autocovariance théorique du terme d'erreur $\{\mathbf{a}_t\}$. Posons $\hat{\mathbf{a}}_t$, $t = 1, \ldots, n$, les résidus. Les autocovariances résiduelles sont donc $\mathbf{C}_{\hat{\mathbf{a}}}(l) = (C_{\hat{\mathbf{a}},ij}(l))_{i,j=1,\ldots,d}$, $l = 1, \ldots, n-1$, où :

$$\mathbf{C}_{\hat{\mathbf{a}}}(l) = \frac{1}{n} \sum_{t=l+1}^{n} \hat{\mathbf{a}}_t \hat{\mathbf{a}}_{t-l}^{\top}.$$

On pose $\mathbf{c}_{\mathbf{a}}(l) = \operatorname{vec}\{\mathbf{C}_{\mathbf{a}}(l)\}$, $\mathbf{c}_{\mathbf{a}} = (\mathbf{c}_{\mathbf{a}}^{\top}(1), \mathbf{c}_{\mathbf{a}}^{\top}(2), ..., \mathbf{c}_{\mathbf{a}}^{\top}(M))^{\top}$, où M est une constante finie à spécifier. L'opérateur $\operatorname{vec}(\cdot)$ empile les colonnes d'une matrice. Il est étudié en particulier dans Harville (1997). Les propriétés des autocovariances et autocorrélations résiduelles ont été étudiées par plusieurs auteurs. Li et McLeod (1981) ont établi la distribution asymptotique normale de $\mathbf{c}_{\hat{\mathbf{a}}}$ dans les modèles VAR et VARMA. Duchesne (2005) donne la distribution asymptotique des autocovariances et autocorrélations résiduelles dans les modèles VAR avec variables exogènes. Voir aussi Lütkepohl (2005) et Li (2004) pour d'autres résultats dans le cas multivarié.

1.5.3. Tests basés sur la matrice d'autocovariance résiduelle pour délais individuels

Les tests basés sur les autocovariances résiduelles sont motivés par le fait que si $\mathbf{Y}_t$, $t = 1, \ldots, n$, est une série chronologique dont les erreurs sont non-corrélées, alors le modèle est adéquat si $\mathbf{\Gamma}_{\mathbf{a}}(l) = \mathbf{0}$, $l = 1, 2, \ldots$. On est donc intéressé à tester si les autocovariances résiduelles sont significativement différentes de $\mathbf{0}$. Pour les tests aux délais individuels, on est intéressé de tester l'hypothèse nulle $\mathbf{\Gamma}_{\mathbf{a}}(l) = \mathbf{0}$, pour un l en particulier. Soit $\hat{\mathbf{\Omega}}(l)$, un estimateur convergeant de $\mathbf{\Omega}(l)$, la matrice de covariance asymptotique de $\mathbf{c}_{\hat{\mathbf{a}}}(l)$. Définissons la statistique de test :

$$H(l) = n\mathbf{c}_{\hat{\mathbf{a}}}(l)^{\top} \hat{\mathbf{\Omega}}(l)^{-1} \mathbf{c}_{\hat{\mathbf{a}}}(l). \tag{1.5.1}$$

Sous l'hypothèse nulle de l'adéquation du modèle, $H(l) \xrightarrow{d} \chi^2_{d^2}$. Les grandes valeurs de $H(l)$ sont celles qui entraînent le rejet de l'hypothèse nulle. Ljung et Box (1978) et Hosking (1980) ont discuté d'une statistique modifiée dont les propriétés empiriques sont plus satisfaisantes. Cette statistique est définie comme suit :

$$Q(l) = \frac{n}{n-l} H(l). \tag{1.5.2}$$

Dans de nombreuses études de simulation, il ressort que pour un n fini, la statistique de test $Q(l)$ affiche des niveaux empiriques plus près des niveaux nominaux que $H(l)$, particulièrement pour des délais l élevés.

1.5.4. Tests de type portemanteau

On peut être intéressé de tester la signification conjointe des $\mathbf{c}_{\mathbf{a}}(l)$, $l = 1, \ldots, M$. Dans ce cas, un test de type portemanteau est plus approprié. Soit $\hat{\boldsymbol{\Omega}}$, un estimateur convergeant de $\boldsymbol{\Omega}$, la matrice de covariance asymptotique de $(\mathbf{c}_{\mathbf{a}}^\top(1), \ldots, \mathbf{c}_{\mathbf{a}}^\top(M))^\top$. Posons :

$$H_M = n\mathbf{c}_{\hat{\mathbf{a}}}^\top \hat{\boldsymbol{\Omega}}^{-1} \mathbf{c}_{\hat{\mathbf{a}}}. \tag{1.5.3}$$

Sous l'hypothèse nulle de l'adéquation du modèle, $H_M \xrightarrow{d} \chi^2_{Md^2}$. Les grandes valeurs de H_M sont encore une fois celles qui entraînent le rejet de l'hypothèse nulle. Comme pour la statistique à délai individuel $H(l)$, les niveaux empiriques de la statistique H_M sont souvent éloignés des niveaux nominaux. Ljung et Box (1978) discutent de la statistique portemanteau modifiée :

$$Q_M = n\mathbf{c}_{\hat{\mathbf{a}}}^{*\top} \hat{\boldsymbol{\Omega}}^{-1} \mathbf{c}_{\hat{\mathbf{a}}}^*, \tag{1.5.4}$$

où $\mathbf{c}_{\mathbf{a}}^* = (\sqrt{n/n-1}\mathbf{c}_{\mathbf{a}}^\top(1), \ldots, \sqrt{n/n-M}\mathbf{c}_{\mathbf{a}}^\top(M))^\top$. En pratique, la statistique Q_M respecte souvent mieux les niveaux que H_M, en particulier pour de grandes valeurs de M.

L'objectif de ce premier chapitre visait à présenter les concepts les plus utiles en séries chronologiques multivariées. Nous avons survolé l'estimation et la validation des modèles multivariés. Dans l'article qui suit, nous développons des tests originaux pour diagnostiquer des modèles non-linéaires multivariés. Ces tests généralisent les résultats univariés de Li (1992), qui a établi la distribution asymptotique des autocovariances et autocorrélations résiduelles pour des modèles non-linéaires univariés. Cependant, Li (1992) supposait que le terme d'erreur de son modèle était un bruit blanc fort. Dans nos développements, nous supposons simplement que le terme d'erreur est une différence de martingale. Comme nous allons le constater, la distribution asymptotique peut être passablement différente comparativement au cas fort. Cet état de fait est illustré dans une étude de simulation où les tests de Li (1992) supposant un bruit blanc fort sont comparés à nos tests, sous différentes conditions pour le terme d'erreur de nos modèles.

Chapitre 2

ARTICLE

Le présent chapitre est une copie intégrale de l'article intitulé "*Diagnostic checking multivariate nonlinear time series models with martingale difference errors*" soumis dans une revue scientifique le 25 mai 2007. Le premier auteur est également l'auteur du présent mémoire.

Diagnostic checking multivariate nonlinear time series models with martingale difference errors[1]

Dominique CHABOT-HALLÉ

Département de mathématiques et statistique

Université de Montréal

and

Pierre DUCHESNE[2]

Département de mathématiques et statistique

Université de Montréal

May 25, 2007

Abstract

In this article, we derive the asymptotic distribution of residual autocovariance and autocorrelation matrices for a general class of multivariate nonlinear time series models assuming only that the error term is a martingale difference sequence. Two types of applications are developed : global test statistics of the portmanteau type and one-lag test statistics, which describe the residual correlation at individual lags. To illustrate the proposed methodology, simulation results are reported for diagnosing multivariate threshold time series models. The following

[1]Abbreviated title : "Diagnostic checking multivariate nonlinear models".

Corresponding author : Pierre Duchesne, Université de Montréal ; Département de mathématiques et de statistique ; C.P. 6128 Succursale Centre-Ville ; Montréal, Québec H3C 3J7 ; Canada. tel : (514) 343-7267 ; fax : (514) 343-5700 ; e-mail : [redacted]

[2]This work was supported by a grant from the Natural Science and Engineering Research Council of Canada.

test statistics are compared : the classical test statistics presuming independent errors and the proposed methodology which supposes only martingale difference errors.

Key words and phrases : Autocovariance matrices ; diagnostic checking ; martingale difference sequence ; multivariate time series ; nonlinear time series ; portmanteau test statistics.
Mathematics subject classification codes (2000) : primary 62M10 ; secondary 62H10.

1. Introduction

Let $\mathbf{Y}_t = \{\mathbf{Y}_t, t \in \mathbb{Z}\}$ be a stationary and ergodic multivariate stochastic process, where $\mathbf{Y}_t = (Y_t(1), \ldots, Y_t(d))^\top$. We consider the following multivariate nonlinear time series model, defined by :

$$\mathbf{Y}_t = \mathbf{f}(\mathcal{I}_{t-1}; \boldsymbol{\theta}_0) + \mathbf{a}_t, \tag{2.1}$$

where $\mathcal{I}_t$ represents $\sigma(\mathbf{Y}_t, \mathbf{Y}_{t-1}, \ldots)$, the sigma-field generated by $\{\mathbf{Y}_t, \mathbf{Y}_{t-1}, \ldots\}$, $\mathbf{f} = (f_1, \ldots, f_d)^\top$ is a known real-valued function with values in $\mathbb{R}^d$, and $\boldsymbol{\theta}_0$ corresponds to a $K \times 1$ vector of unknown parameters. It is assumed that each $f_i \equiv f_i(\mathcal{I}_{t-1}; \boldsymbol{\theta})$, $i = 1, \ldots, d$, is function of the past $\mathbf{Y}_t$'s and of $\boldsymbol{\theta}$. The vector function $\mathbf{f}(\cdot; \cdot)$ is supposed to have continuous second order derivatives with respect to $\boldsymbol{\theta}$, almost surely. The error term $\mathbf{a} = \{\mathbf{a}_t, t \in \mathbb{Z}\}$ is a difference martingale sequence, that is $E(\mathbf{a}_t | \, \mathcal{I}_{t-1}) = \mathbf{0}$ almost surely, where $\mathbf{a}_t = (a_t(1), \ldots, a_t(d))^\top$, $t \in \mathbb{Z}$. This implies that $\mathbf{a}$ forms an uncorrelated but not necessarily independent sequence of random vectors with mean zero. We denote $\boldsymbol{\Sigma}_{\mathbf{a}} = (\sigma_{\mathbf{a},ij})_{i,j=1,\ldots,d}$ the covariance matrix of $\mathbf{a}_t$, that we assume non-singular. For example, the error term $\mathbf{a}$ could belong to the general class of multivariate (generalized) autoregressive conditional heteroskedastic [(G)ARCH] processes. See Gouriéroux (1997) for an introduction to these models. It is further supposed that (2.1) is an invertible stochastic process ; equivalently, this means that the process $\mathbf{a}$ is assumed measurable with respect to $\mathcal{I}_t$.

The nonlinear model (2.1) represents a very general class of time series models with a general specification for the error term. With the emergence of economic, financial and other complex applications, nonlinear time series have generated considerable interest in the last years. Many classes of models have been developed for univariate time series models. The pioneer work of Granger and Andersen (1978) proposed the class of bilinear models, while Tong (1978, 1983) and Tong and Lim (1980) introduced and studied threshold models, such as the class of threshold autoregressive models (TAR) and self-exciting TAR models (SETAR). Extending the idea of using probability switching, but emphasizing aperiodic transition between various states, led Hamilton (1989) to propose Markov switching models. The thresholds in TAR models correspond to discontinuity points in the conditional mean function. Smooth TAR models are obtained by using smooth transition functions : the resulting models form the class of smooth transition autoregressive models (STAR). See Chan and Tong (1986). Teräsvirta (1994) investigated modelling economic and financial applications using STAR models. Tong (1990), Granger and Teräsvirta (1993) and Tsay (2005) provide surveys of univariate nonlinear models. In multivariate models, a general class of threshold models is discussed in Tsay (1998) and van Dick, Teräsvirta and Franses (2002) present some natural extensions of STAR models for the multivariate framework.

Let $\mathbf{Y}_t$, $t = 1, \ldots, n$ be a finite realization of process $\mathbf{Y}$. An important aspect in any statistical modelling application appears to validate a particular adjusted model such as (2.1), following a certain estimating procedure (conditional least squares, maximum likelihood estimation, etc.). In our set-up, this means checking if the conditional mean function $\mathbf{f}(\cdot;\cdot)$ seems correctly specified; no residual correlation should be statistically significant from the fitted model. Under precise regularity conditions, allowing for the uncorrelated error term to be not necessarily Gaussian, Klimko and Nelson (1978) considered the properties of conditional least squares estimators in univariate nonlinear time series models. Tjøstheim (1986) generalized the results of Klimko and Nelson (1978) for general multivariate nonlinear time series models, giving conditions for strong consistency and asymptotic normality using an arbitrarily penalty function, and specializing the results for conditional least squares and maximum likelihood type estimators. See also Taniguchi and Kakizawa (2000, Section 3.2). Note that, typically, after adjusting a particular model, if the residuals seem uncorrelated but dependent, a conditional variance function could be specified, such as a multivariate (G)ARCH model, assuming the conditional mean function $\mathbf{f}(\cdot;\cdot)$ correctly specified. This conditional variance function could be validated using test statistics such as the ones proposed by Ling and Li (1997) or Duchesne (2004), amongst others.

Residual autocovariances and autocorrelations have been found appropriate tools for diagnosing many time series models. The monograph of Li (2004) presents an overview of diagnostic checking for univariate and multivariate time series models. In univariate nonlinear time series models, Li (1992) provides the asymptotic distribution of the residual autocorrelations for some very general nonlinear time series models, assuming Gaussian errors. Illustrating his results for univariate threshold models, he notes that the common value $n^{-1/2}$, obtained by neglecting parameter estimation uncertainty, represents only a rough guide in diagnostic checking ; in fact, the use of $n^{-1/2}$ can be very conservative, especially for the first few lags. Hwang, Basawa and Reeves (1994) extended the work of Li (1992) to include linear and nonlinear models with random parameters, assuming independent errors. For univariate linear time series models with general error terms, Francq, Roy and Zakoïan (2005) consider diagnostic checking autoregressive-moving-average (ARMA) models with uncorrelated errors. Recently, Francq and Raïssi (2007) have extended the approach of Francq, Roy and Zakoïan (2005) to vector autoregressive (VAR) models with uncorrelated but nonindependent errors.

From a practical point of view, it appears important to describe precisely the serial correlation among the residuals, taking into account the possible interaction between the time series ; the univariate result of Li (1992) appears of limited use if the d univariate time series composing (2.1) are not independent. Furthermore, it is generally recognized that financial data exhibit heavier tails than those coming from the normal distribution, and they seem often compatible with (G)ARCH mechanisms. These phenomena are not the exceptions, and seem established facts in economic and financial applications (see, e.g., Tsay (2005)). Consequently, it appears important to relax the assumption of independent and Gaussian errors. In view of this, the aim of this paper is to extend the univariate results of Li (1992) in two important directions : (i) the more realistic situation of multivariate data ; (ii) the error term is supposed to be a martingale difference sequence. Our theorem demonstrates that the asymptotic variance of the residual autocovariances and autocorrelations may be very different than the one obtained by Li (1992), which assumes independent errors. Interestingly, it may be estimated as simply as in the independent case. As an application of our main result, two types of test statistics are proposed. In the first category, portmanteau test statistics are obtained. Because of the non-linear structure of the conditional mean function and of the covariance between the residual autocovariances, the proposed portmanteau test statistics are different than Hosking's (1980) multivariate portmanteau test statistic, which is valid for vector ARMA models with independent errors. Test statistics measuring the dependence at individual lags are also proposed, relying on the lag-l residual autocovariance function ; they fall in the second category. We give an explicit formula for the asymptotic covariance structure of the residual autocovariances, and as Li (1992), this asymptotic expression is consistently estimated in our testing methodology.

The paper is organized as follows. In Section 2, we derive the asymptotic distribution of a vector of lag-l residual autocovariances matrices, $l = 1, \dots, M$, using an approach similar to the one of Li (1992), but adapted for multivariate time series, and assuming only a martingale difference white noise. We describe in Section 3 some applications to diagnostic checking : portmanteau test statistics and one-lag test statistics. Finally, some simulation experiments are conducted in Section 4 for diagnosing multivariate threshold time series models. The following test statistics are compared : the classical test statistics presuming independent errors and the proposed methodology which assumes only martingale difference error terms. Section 5 concludes.

2. Asymptotic distribution of residual autocovariance matrices

The conditional least squares estimators $\hat{\boldsymbol{\theta}}_n$ of $\boldsymbol{\theta}_0$ is obtained by minimizing the criterion :

$$S_n(\boldsymbol{\theta}) = \sum_{t=\tau+1}^{n} \{\mathbf{Y}_t - \mathbf{f}(\mathcal{I}_{t-1};\boldsymbol{\theta})\}^\top \boldsymbol{\Sigma}_{\mathbf{a}}^{-1} \{\mathbf{Y}_t - \mathbf{f}(\mathcal{I}_{t-1};\boldsymbol{\theta})\},$$

where τ is an appropriate integer. For example, if $\{\mathbf{Y}_t\}$ represents a nonlinear autoregressive process of order p, a possible choice of τ is $\tau = p$ (Taniguchi and Kakizawa, 2000, p. 98). Under some regularity conditions, see Tjøstheim (1986), it can be shown that $\hat{\boldsymbol{\theta}}_n$ is asymptotically normal. In the notation of Serfling (1980) :

$$\hat{\boldsymbol{\theta}}_n \text{ is } AN\left(\boldsymbol{\theta}_0, \frac{\mathbf{U}^{-1}\mathbf{R}\mathbf{U}^{-1}}{n}\right),$$

where

$$\mathbf{U} = E\left\{\frac{\partial(\mathbf{Y}_t - \mathbf{f}_{t-1})^\top}{\partial\boldsymbol{\theta}}\boldsymbol{\Sigma}_{\mathbf{a}}^{-1}\frac{\partial(\mathbf{Y}_t - \mathbf{f}_{t-1})}{\partial\boldsymbol{\theta}^\top}\right\}, \tag{2.2}$$

$$\mathbf{R} = E\left\{\frac{\partial\mathbf{f}_{t-1}^\top}{\partial\boldsymbol{\theta}}\boldsymbol{\Sigma}_{\mathbf{a}}^{-1}(\mathbf{Y}_t - \mathbf{f}_{t-1})(\mathbf{Y}_t - \mathbf{f}_{t-1})^\top\boldsymbol{\Sigma}_{\mathbf{a}}^{-1}\frac{\partial\mathbf{f}_{t-1}}{\partial\boldsymbol{\theta}^\top}\right\}, \tag{2.3}$$

with $\mathbf{f}_{t-1} = \mathbf{f}(\mathcal{I}_{t-1};\boldsymbol{\theta}_0)$. Let $\boldsymbol{\Gamma}_{\mathbf{a}}(l) = \mathrm{cov}(\mathbf{a}_t, \mathbf{a}_{t-l})$ be the lag-l theoretical autocovariance, and $\boldsymbol{\Gamma}_0 = \mathrm{diag}(\sigma_{\mathbf{a},11}^{1/2}, \ldots, \sigma_{\mathbf{a},dd}^{1/2})$. We define $\boldsymbol{\rho}_{\mathbf{a}}(l) = \boldsymbol{\Gamma}_0^{-1}\boldsymbol{\Gamma}_{\mathbf{a}}(l)\boldsymbol{\Gamma}_0^{-1}$ the lag-l theoretical autocorrelation. The sample autocovariance matrices satisfy $\mathbf{C}_{\mathbf{a}}(l) = (C_{\mathbf{a},ij}(l))_{i,j=1,\ldots,d}$ and they are given by :

$$\mathbf{C}_{\mathbf{a}}(l) = \begin{cases} n^{-1}\sum_{t=l+1}^{n}\mathbf{a}_t\mathbf{a}_{t-l}^\top, & l = 0, 1, \ldots, n-1, \\ \mathbf{C}_{\mathbf{a}}^\top(-l), & l = -1, \ldots, -n+1. \end{cases}$$

Let $\mathbf{c}_{\mathbf{a}} = (\mathbf{c}_{\mathbf{a}}^\top(1), \ldots, \mathbf{c}_{\mathbf{a}}^\top(M))^\top$, $\mathbf{c}_{\mathbf{a}}(l) = \mathrm{vec}\{\mathbf{C}_{\mathbf{a}}(l)\}$ be the vector of sample autocovariances and consider $\mathbf{D}_{\mathbf{a}} = \mathrm{diag}(C_{\mathbf{a},11}^{1/2}(0), \ldots, C_{\mathbf{a},dd}^{1/2}(0))$, where $\mathrm{vec}(\mathbf{A})$ denotes the vector obtained by stacking the columns of matrix $\mathbf{A}$ (see Harville, 1997, Chapter 16.3). We define similarly the vectors of sample autocorrelations $\mathbf{r}_{\mathbf{a}} = (\mathbf{r}_{\mathbf{a}}^\top(1), \ldots, \mathbf{r}_{\mathbf{a}}^\top(M))^\top$ and $\mathbf{r}_{\mathbf{a}}(l) = \mathrm{vec}\{\mathbf{D}_{\mathbf{a}}^{-1}\mathbf{C}_{\mathbf{a}}(l)\mathbf{D}_{\mathbf{a}}^{-1}\} = (\mathbf{D}_{\mathbf{a}}^{-1} \otimes \mathbf{D}_{\mathbf{a}}^{-1})\,\mathbf{c}_{\mathbf{a}}(l)$, where '$\otimes$' denotes the Kronecker product (see Harville, 1997, Chapter 16.1). Here M denotes a fixed integer (with respect to the sample size n), satisfying $1 \leq M < n$, which corresponds to the maximal lag order.

Let $\hat{\mathbf{f}}_{t-1} = \mathbf{f}(\mathcal{I}_{t-1};\hat{\boldsymbol{\theta}}_n)$ and introduce the model residuals $\hat{\mathbf{a}}_t = \mathbf{Y}_t - \hat{\mathbf{f}}_{t-1}$. The vectors of residual autocovariances and autocorrelations, $\mathbf{c}_{\hat{\mathbf{a}}}$, $\mathbf{c}_{\hat{\mathbf{a}}}(l)$, $|l| < n$, $\mathbf{r}_{\hat{\mathbf{a}}}$ and $\mathbf{r}_{\hat{\mathbf{a}}}(l)$, $|l| < n$, are defined naturally. We now derive the asymptotic distribution of the vector $\mathbf{c}_{\hat{\mathbf{a}}}$. Under our assumptions for $\{\mathbf{a}_t\}$, it can be shown that $n^{1/2}\mathbf{c}_{\mathbf{a}}$ converges in distribution to a normally distributed random vector :

$$\mathbf{c}_{\mathbf{a}} \text{ is } AN\left(\mathbf{0}, \frac{\boldsymbol{\Delta}_M}{n}\right),$$

where

$$\boldsymbol{\Delta}_M = (\boldsymbol{\Delta}_{ij})_{i,j=1,\ldots,M}, \tag{2.4}$$

with $\boldsymbol{\Delta}_{ij} = E(\mathbf{a}_{t-i}\mathbf{a}_{t-j}^\top \otimes \mathbf{a}_t\mathbf{a}_t^\top)$, $i, j = 1, \ldots, M$. In particular, when the error term corresponds to a strong white noise, that is the $\mathbf{a}_t$'s form an independent sequence of random vectors with mean zero and variance $\boldsymbol{\Sigma}_{\mathbf{a}}$, then $\boldsymbol{\Delta}_{ll} = \boldsymbol{\Sigma}_{\mathbf{a}} \otimes \boldsymbol{\Sigma}_{\mathbf{a}}$, $\boldsymbol{\Delta}_{lm} = \mathbf{0}$,

$l \neq m$, for $l, m \in \{1, \ldots, M\}$, and we deduce that the asymptotic variance of $\mathbf{c_a}$ reduces to $\mathbf{\Delta}_M = \mathbf{I}_M \otimes \mathbf{\Sigma_a} \otimes \mathbf{\Sigma_a}$, where $\mathbf{I}_M$ denotes the $M \times M$ identity matrix ; we retrieve a result due to Li and McLeod (1981). We give explicit expressions for $\partial \mathbf{c_a}(l)/\partial \boldsymbol{\theta}^\top$, $l = 1, \ldots, M$. Using a standard matrix differentiation result and the martingale difference property of $\{\mathbf{a}_t\}$, it follows that :

$$\begin{aligned}
\frac{\partial \mathbf{c_a}(l)}{\partial \boldsymbol{\theta}^\top} &= n^{-1} \sum_{t=l+1}^{n} \frac{\partial \mathrm{vec}(\{\mathbf{Y}_t - \mathbf{f}_{t-1}\}\{\mathbf{Y}_{t-l} - \mathbf{f}_{t-l-1}\}^\top)}{\partial \boldsymbol{\theta}^\top}, \\
&= n^{-1} \sum_{t=l+1}^{n} (\mathbf{I}_d \otimes \{\mathbf{Y}_t - \mathbf{f}_{t-1}\}) \frac{\partial \{\mathbf{Y}_{t-l} - \mathbf{f}_{t-l-1}\}}{\partial \boldsymbol{\theta}^\top} \\
&\quad + (\{\mathbf{Y}_{t-l} - \mathbf{f}_{t-l-1}\} \otimes \mathbf{I}_d) \frac{\partial \{\mathbf{Y}_t - \mathbf{f}_{t-1}\}}{\partial \boldsymbol{\theta}^\top}, \\
&\xrightarrow{p} -\mathbf{J}_l,
\end{aligned}$$

where $\mathbf{J}_l = E[\{\mathbf{Y}_{t-l} - \mathbf{f}_{t-l-1}\} \otimes \partial \mathbf{f}_{t-1}/\partial \boldsymbol{\theta}^\top]$. By expanding $\mathbf{c}_{\hat{\mathbf{a}}}$ in a Taylor series expansion, we obtain that :

$$\mathbf{c}_{\hat{\mathbf{a}}} = \mathbf{c_a} + \frac{\partial \mathbf{c_a}}{\partial \boldsymbol{\theta}^\top} \left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 \right) + \mathbf{o}_P(n^{-1/2}),$$

where the derivative $\partial \mathbf{c_a}/\partial \boldsymbol{\theta}^\top$ represents a $Md^2 \times K$ matrix satisfying the relation $\left(\partial \mathbf{c_a}/\partial \boldsymbol{\theta}^\top\right)^\top = \left((\partial \mathbf{c_a}(1)/\partial \boldsymbol{\theta}^\top)^\top, \ldots, (\partial \mathbf{c_a}(M)/\partial \boldsymbol{\theta}^\top)^\top\right)$. Thus $\mathbf{c}_{\hat{\mathbf{a}}} = \mathbf{c_a} - \mathbf{J}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + \mathbf{o}_P(n^{-1/2})$, where :

$$\mathbf{J}^\top = (\mathbf{J}_1^\top, \ldots, \mathbf{J}_M^\top). \tag{2.5}$$

Tjøstheim (1986) has shown that :

$$n^{1/2}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = -\frac{1}{2}\mathbf{U}^{-1} n^{-1/2} \frac{\partial S_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} + \mathbf{o}_P(1),$$

where $\mathbf{U} = (2n)^{-1} \partial^2 S_n(\boldsymbol{\theta}_0)/\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top$. See also Taniguchi and Kakizawa (2000, pp. 98-100). Furthermore, the asymptotic distribution of $n^{-1/2} \partial S_n(\boldsymbol{\theta}_0)/\partial \boldsymbol{\theta}$ is normal and the joint asymptotic normality of $n^{1/2}\left((\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)^\top, \mathbf{c}_\mathbf{a}^\top\right)$ can be established. It follows that the asymptotic covariance structure between $n^{1/2}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ and $n^{1/2}\mathbf{c_a}$ is given by :

$$\lim_n n\mathrm{cov}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0, \mathbf{c_a}) = -\frac{1}{2}\mathbf{U}^{-1} \lim_n E\left\{ \frac{\partial S_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \mathbf{c}_\mathbf{a}^\top \right\}.$$

Using straightforward matrix differentiation :

$$\frac{\partial S_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}^\top} = -2 \sum_{t=\tau+1}^{n} \left[\{\mathbf{Y}_t - \mathbf{f}(\mathcal{I}_{t-1}, \boldsymbol{\theta}_0)\}^\top \boldsymbol{\Sigma}_{\mathbf{a}}^{-1} \frac{\partial \mathbf{f}_{t-1}}{\partial \boldsymbol{\theta}^\top} \right].$$

We now derive an explicit expression for the matrix $E\{\partial S_n(\boldsymbol{\theta}_0)/\partial \boldsymbol{\theta} \mathbf{c}_{\mathbf{a}}^\top\}$. Note that :

$$\frac{\partial S_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \mathbf{c}_{\mathbf{a}}^\top = \left(\frac{\partial S_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \mathbf{c}_{\mathbf{a}}^\top(1), \ldots, \frac{\partial S_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \mathbf{c}_{\mathbf{a}}^\top(M) \right).$$

We evaluate $E[\partial S_n(\boldsymbol{\theta}_0)/\partial \boldsymbol{\theta} \text{vec}^\top\{\mathbf{C}_{\mathbf{a}}(j)\}]$. Since

$$\frac{\partial S_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \text{vec}^\top \{\mathbf{C}_{\mathbf{a}}(j)\} = -\frac{2}{n} \sum_{t=\tau+1}^{n} \sum_{s=j+1}^{n} \left(\frac{\partial \mathbf{f}_{t-1}^\top}{\partial \boldsymbol{\theta}} \boldsymbol{\Sigma}_{\mathbf{a}}^{-1} \mathbf{a}_t \right) (\mathbf{a}_{s-j}^\top \otimes \mathbf{a}_s^\top),$$

the martingale difference property of $\{\mathbf{a}_t\}$ implies :

$$E\left[\frac{\partial S_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \text{vec}^\top \{\mathbf{C}_{\mathbf{a}}(j)\} \right] = -\frac{2}{n} \sum_{t=\tau+1}^{n} E\left(\mathbf{a}_{t-j}^\top \otimes \frac{\partial \mathbf{f}_{t-1}^\top}{\partial \boldsymbol{\theta}} \boldsymbol{\Sigma}_{\mathbf{a}}^{-1} \mathbf{a}_t \mathbf{a}_t^\top \right) \to -2\mathbf{J}_j^{*\top},$$

where $\mathbf{J}_j^* = E[\{\mathbf{Y}_{t-j} - \mathbf{f}_{t-j-1}\} \otimes \mathbf{a}_t \mathbf{a}_t^\top \boldsymbol{\Sigma}_{\mathbf{a}}^{-1} \partial \mathbf{f}_{t-1}/\partial \boldsymbol{\theta}^\top]$. Thus $-\frac{1}{2}E[\partial S_n(\boldsymbol{\theta}_0)/\partial \boldsymbol{\theta} \mathbf{c}_{\mathbf{a}}^\top] \to \mathbf{J}^{*\top}$ and $\lim_n n\text{cov}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0, \mathbf{c}_{\mathbf{a}}) = \mathbf{U}^{-1}\mathbf{J}^{*\top}$, where

$$\mathbf{J}^{*\top} = (\mathbf{J}_1^{*\top}, \ldots, \mathbf{J}_M^{*\top}). \tag{2.6}$$

This leads us to our main theorem.

Théorème 2.0.1. *Let a time series be generated by equation (2.1). Let $\hat{\boldsymbol{\theta}}_n$ be the conditional least square estimator of $\boldsymbol{\theta}_0$. Then*

$$\mathbf{c}_{\hat{\mathbf{a}}} \text{ is } AN\left(\mathbf{0}, \frac{\boldsymbol{\Omega}}{n}\right),$$

where $\boldsymbol{\Omega} = \boldsymbol{\Delta}_M - \mathbf{J}^*\mathbf{U}^{-1}\mathbf{J}^\top - \mathbf{J}\mathbf{U}^{-1}\mathbf{J}^{*\top} + \mathbf{J}\mathbf{U}^{-1}\mathbf{R}\mathbf{U}^{-1}\mathbf{J}^\top$, and $\mathbf{U}$, $\mathbf{R}$, $\boldsymbol{\Delta}_M$, $\mathbf{J}$ and $\mathbf{J}^*$ are defined by (2.2), (2.3), (2.4), (2.5) and (2.6), respectively.

An immediate corollary of Theorem 2.0.1 gives the asymptotic distribution of the residual autocorrelations. Let $\mathbf{L} = \mathbf{I}_M \otimes \boldsymbol{\Gamma}_0^{-1} \otimes \boldsymbol{\Gamma}_0^{-1}$. Since $\mathbf{D}_{\hat{\mathbf{a}}}$ converges to $\boldsymbol{\Gamma}_0$ in probability, that is $\mathbf{D}_{\hat{\mathbf{a}}} \xrightarrow{p} \boldsymbol{\Gamma}_0$, the asymptotic distribution of $n^{1/2}\mathbf{r}_{\hat{\mathbf{a}}}$ is obtained by an appropriate scaling based on Theorem 2.0.1 :

$$\mathbf{r}_{\hat{\mathbf{a}}} \text{ is } AN\left(\mathbf{0}, \frac{\mathbf{L}\boldsymbol{\Omega}\mathbf{L}^\top}{n}\right).$$

If $\{\mathbf{a}_t\}$ is a strong white noise, it follows that $\mathbf{J}^* = \mathbf{J}$, $\mathbf{R} = \mathbf{U}$ and thus the covariance structure of the residual autocovariances $\mathbf{c}_{\hat{\mathbf{a}}}(l)$, $l = 1, \ldots, M$, and $\mathbf{c}_{\hat{\mathbf{a}}}$

reduce to the familiar formulas $\boldsymbol{\Omega}_{Sl} = \boldsymbol{\Sigma}_{\mathbf{a}} \otimes \boldsymbol{\Sigma}_{\mathbf{a}} - \mathbf{J}_l \mathbf{U}^{-1} \mathbf{J}_l^{\top}$, $l = 1, \ldots, M$, and $\boldsymbol{\Omega}_S = \mathbf{I}_M \otimes \boldsymbol{\Sigma}_{\mathbf{a}} \otimes \boldsymbol{\Sigma}_{\mathbf{a}} - \mathbf{J}\mathbf{U}^{-1}\mathbf{J}^{\top}$, respectively. A similar result holds for the residual autocorrelations $\mathbf{r}_{\hat{\mathbf{a}}}(l)$, $l = 1, \ldots, M$, and $\mathbf{r}_{\hat{\mathbf{a}}}$. For univariate nonlinear time series, letting $d = 1$ and assuming independence of the a_t's allow us to retrieve a Theorem due to Li (1992, p. 436).

In view of Theorem 2.0.1, it appears that if the error term represents a martingale difference sequence, the asymptotic variance is different than the one obtained assuming independent errors. This suggests that the familiar formulas for the asymptotic variance of the residual autocovariances and autocorrelations, and the classical portmanteau test statistics, could lead to incorrect inference when diagnosing multivariate nonlinear time series with uncorrelated but dependent errors. This is illustrated in Section 4.

3. Applications for diagnostic checking multivariate nonlinear models

Theorem 2.0.1 gives the asymptotic distribution of the autocovariance matrices $\mathbf{C}_{\hat{\mathbf{a}}}(l)$, $l = 1, \ldots, M$. This result allows us to construct test statistics for checking the adequacy of multivariate nonlinear time series models. Two types of test statistics are proposed. The first one, of the portmanteau type, integrates several lags and is based on $\mathbf{C}_{\hat{\mathbf{a}}}(l)$, $l = 1, \ldots, M$, whilst the second one relies, for a particular lag l, say, on the covariance matrix $\mathbf{C}_{\hat{\mathbf{a}}}(l)$.

3.1 Portmanteau test statistic

The null hypothesis of model adequacy is $\boldsymbol{\Gamma}_{\mathbf{a}}(l) = \mathbf{0}$, $l = 1, 2, \ldots$. Let

$$\hat{\boldsymbol{\Omega}} = \hat{\boldsymbol{\Delta}}_M - \hat{\mathbf{J}}^{*}\hat{\mathbf{U}}^{-1}\hat{\mathbf{J}}^{\top} - \hat{\mathbf{J}}\hat{\mathbf{U}}^{-1}\hat{\mathbf{J}}^{*\top} + \hat{\mathbf{J}}\hat{\mathbf{U}}^{-1}\hat{\mathbf{R}}\hat{\mathbf{U}}^{-1}\hat{\mathbf{J}}^{\top}$$

be a consistent estimator of $\boldsymbol{\Omega}$. Consequently, $n\mathbf{c}_{\hat{\mathbf{a}}}^{\top}\hat{\boldsymbol{\Omega}}^{-1}\mathbf{c}_{\hat{\mathbf{a}}} \xrightarrow{d} \chi^2_{d^2M}$ under the null hypothesis; this test statistic can be used for testing the joint significance of $\mathbf{c}_{\hat{\mathbf{a}}}(l)$, $l = 1, \ldots, M$. Note that as Box-Pierce type test statistics, the asymptotic χ^2 distribution may be poorly approximated for large M (see Li (2004), amongst others). Following Ljung and Box (1978), we introduce a level-adjusted portmanteau test statistic in order to check the adequacy of a particular nonlinear model :

$$Q_M = n\mathbf{c}_{\hat{\mathbf{a}}}^{*\top}\hat{\mathbf{\Omega}}^{-1}\mathbf{c}_{\hat{\mathbf{a}}}^{*}, \tag{2.7}$$

where $\mathbf{c}_{\hat{\mathbf{a}}}^{*} = (\sqrt{n/(n-1)}\mathbf{c}_{\hat{\mathbf{a}}}^{\top}(1), \ldots, \sqrt{n/(n-M)}\mathbf{c}_{\hat{\mathbf{a}}}^{\top}(M))^{\top}$. In many practical situations, the level-adjusted test statistic (2.7) should display better empirical properties than to use a test statistic based on $\mathbf{c}_{\hat{\mathbf{a}}}$. The procedure rejects the null hypothesis if Q_M is large, more precisely if $Q_M > \chi^2_{Md^2,1-\alpha}$, where $\chi^2_{Md^2,1-\alpha}$ denotes the $1-\alpha$ quantile of the $\chi^2_{Md^2}$ distribution.

3.2 One-lag test statistics

Portmanteau procedures are useful omnibus measures of adjustment of time series models. However, more insight is gained by considering test statistics describing the dependence at each individual lag l. Let

$$\hat{\mathbf{\Omega}}_l = \hat{\mathbf{\Delta}}_{ll} - \hat{\mathbf{J}}_l^{*}\hat{\mathbf{U}}^{-1}\hat{\mathbf{J}}_l^{\top} - \hat{\mathbf{J}}_l\hat{\mathbf{U}}^{-1}\hat{\mathbf{J}}_l^{*\top} + \hat{\mathbf{J}}_l\hat{\mathbf{U}}^{-1}\hat{\mathbf{R}}\hat{\mathbf{U}}^{-1}\hat{\mathbf{J}}_l^{\top}$$

be a consistent estimator of $\mathbf{\Omega}_l$, where $\mathbf{\Omega}_l$ represents the asymptotic covariance matrix of $\mathbf{c}_{\hat{\mathbf{a}}}(l)$. We define the following level-adjusted one-lag test statistic :

$$Q(l) = \frac{n^2}{n-l}\mathbf{c}_{\hat{\mathbf{a}}}^{\top}(l)\hat{\mathbf{\Omega}}_l^{-1}\mathbf{c}_{\hat{\mathbf{a}}}(l). \tag{2.8}$$

If the null hypothesis is satisfied, it follows that $Q(l) \xrightarrow{d} \chi^2_{d^2}$. As for the portmanteau procedure Q_M, the one-lag procedure rejects the null hypothesis if $Q(l)$ is too large. As discussed in Hosking (1980) and Duchesne (2005) in vector ARMA models and VAR models with exogenous variables, the level-adjusted test statistic (2.8) may have better finite sample properties than to replace the factor $n^2/(n-l)$ by n in the definition of the test statistic, specially for higher order lags.

If one may want to consider several lags simultaneously, Bonferroni-type adjustments can be performed if one decides to use the test statistics $Q(l)$, $l \in \{1, \ldots, M\}$. This kind of procedure, used in conjunction with the portmanteau test statistic Q_M, contributes to identify the most significant lags.

4. Simulation experiments

In the previous section we introduced portmanteau and one-lag test procedures which should prove useful when the error terms are not independent but martingale difference sequences. The comparison between the proposed test statistics and the test statistics derived under the assumption of independent errors seems to be of particular interest. More precisely, let $\hat{\boldsymbol{\Omega}}_S$ and $\hat{\boldsymbol{\Omega}}_{Sl}$, $l = 1, \ldots, M$, be consistent estimators of $\boldsymbol{\Omega}_S$ and $\boldsymbol{\Omega}_{Sl}$, $l = 1, \ldots, M$, respectively, assuming independent errors. The classical test statistics under independent errors are given by :

$$S_M = n\mathbf{c}_{\hat{\mathbf{a}}}^{*\top}\hat{\boldsymbol{\Omega}}_S^{-1}\mathbf{c}_{\hat{\mathbf{a}}}^{*}, \tag{2.9}$$

$$S(l) = \frac{n^2}{n-l}\mathbf{c}_{\hat{\mathbf{a}}}^{\top}(l)\hat{\boldsymbol{\Omega}}_{Sl}^{-1}\mathbf{c}_{\hat{\mathbf{a}}}(l). \tag{2.10}$$

The aim of our simulation experiments is to compare the exact distributions of the test statistics Q_M, S_M, $Q(l)$, and $S(l)$, with their corresponding asymptotic χ^2 distributions. To undertake this task, we consider the following bivariate threshold model :

$$\mathbf{Y}_t = \begin{cases} \boldsymbol{\Phi}_1^{(1)}\mathbf{Y}_{t-1} + \mathbf{a}_t, & Y_{t-1}(1) < 0, \\ \boldsymbol{\Phi}_1^{(2)}\mathbf{Y}_{t-1} + \mathbf{a}_t, & Y_{t-1}(1) \geq 0, \end{cases} \tag{2.11}$$

where

$$\boldsymbol{\Phi}_1^{(1)} = \begin{pmatrix} 0.7 & 0.0 \\ 0.3 & 0.7 \end{pmatrix}, \quad \boldsymbol{\Phi}_1^{(2)} = \begin{pmatrix} -0.7 & 0.0 \\ -0.3 & -0.7 \end{pmatrix}.$$

The mean function of the process $\{\mathbf{Y}_t\}$ is inspired by a multivariate threshold model studied by Tsay (1998). We considered three processes for the noise $\mathbf{a} = \{\mathbf{a}_t, t \in \mathbb{Z}\}$. First, we included in our study a Gaussian white noise process $\mathbf{a}_1 = \{\mathbf{a}_t\}$, such that $\mathbf{a}_t \sim N_2(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{a}})$, with :

$$\boldsymbol{\Sigma}_{\mathbf{a}} = \begin{pmatrix} 1.0 & 0.2 \\ 0.2 & 1.0 \end{pmatrix}.$$

Second, we investigated $\mathbf{a}_2 = \{\mathbf{a}_t\}$ with $\mathbf{a}_t = \boldsymbol{\eta}_t\eta_{t-1}(1)$, where $\boldsymbol{\eta} = \{\boldsymbol{\eta}_t, t \in \mathbb{Z}\}$, $\boldsymbol{\eta}_t = (\eta_t(1), \eta_t(2))^\top$, is Gaussian white noise, such that $\boldsymbol{\eta}_t \sim N_2(\mathbf{0}, \mathbf{I}_2)$. The white noise process $\mathbf{a}_2$ is composed of uncorrelated but one-dependent random vectors. The error term $\mathbf{a}_2$ is inspired of Romano and Thombs (1996, Exemple 2.1). See also Francq and Raïssi (2007). Finally, we included the ARCH error process $\mathbf{a}_3 = \{\mathbf{a}_t\}$, where $\mathbf{a}_t = \mathbf{H}_t\boldsymbol{\eta}_t$, $\mathbf{H}_t = \text{diag}(H_t(1), H_t(2))$, $H_t(1) = \{0.7 + 0.3a_{t-1}^2(1)\}^{1/2}$, $H_t(2) = \{0.5 + 0.5a_{t-1}^2(2)\}^{1/2}$. The white noise $\mathbf{a}_3$ is conditionally Gaussian with conditional variance matrix $\mathbf{H}_t$.

We examined the empirical frequencies of rejection of the null hypothesis of model adequacy at two different nominal levels (5 and 10 percent) for each of two series lengths ($n = 600, 1200$). For each series length, 1000 independent realizations were generated. For each realization, the true model was estimated by conditional least squares.

For each residual time series, the portmanteau test statistics Q_M and S_M were computed for $M = 2, 3, 6$ and 10. The one-lag test statistics $Q(l)$ and $S(l)$ were computed for $l = 1, \ldots, 10$ but we report the results for $l = 1, \ldots, 5$ since the results for higher lags were very similar.

For each nominal level and for each series of length n, we obtained from the 1000 realizations the empirical frequencies of rejection of the null hypothesis of adequacy. The standard errors of the empirical levels based on 1000 realizations are 0.69% and 0.95% for the nominal levels 5% and 10%, respectively.

The empirical levels of portmanteau test statistics are displayed in Table 1. Under a Gaussian error term, the χ^2 distribution of the test statistics Q_M and S_M gave a satisfactory approximation for all values of M and at the two significance levels, when the series length was $n = 600$. The test statistic S_M admitted very satisfactory empirical levels for all M's considered. This was expected, since the noise $\mathbf{a}_1$ is strong white noise and S_M is justified under the assumption of independent errors. The test statistic Q_M offered a satisfactory behavior and the differences between Q_M and S_M were small in general. As the sample size increased some improvements were generally observed, as expected. Under uncorrelated but dependent white noises, it seemed that the asymptotic distributions for the test statistics Q_M were well approximated by a χ^2 distribution, but in general the χ^2 approximation appeared rather inadequate for S_M. More precisely, under uncorrelated but one-dependent errors the test statistics S_M overrejected the null hypothesis severely, particularly for small values of M. On the other hand, Q_M

TAB. 2.1. Empirical levels (in percentage) of the portmanteau test statistics Q_M and S_M defined by (2.7) and (2.9), for the threshold model given by (2.11) with Gaussian errors, uncorrelated but one-dependent errors and ARCH errors.

	$\alpha = 0.05$				$\alpha = 0.10$			
	$n = 600$		$n = 1200$		$n = 600$		$n = 1200$	
M	Q_M	S_M	Q_M	S_M	Q_M	S_M	Q_M	S_M
	Gaussian errors							
2	4.5	4.2	5.1	5.0	11.5	9.6	10.4	9.2
3	4.5	4.6	5.0	5.5	10.5	9.5	10.6	9.7
6	4.5	5.3	4.8	5.7	11.6	10.4	10.4	11.3
10	4.1	6.1	4.7	5.3	9.7	10.7	10.6	10.8
	Uncorrelated but one-dependent errors							
2	5.3	11.3	6.1	12.0	10.5	18.3	11.8	19.2
3	4.9	9.7	6.1	10.3	10.5	16.9	10.8	15.9
6	5.1	8.7	5.5	10.2	10.5	16.0	11.9	17.5
10	4.7	8.0	4.4	10.1	9.7	14.6	11.0	16.8
	ARCH errors							
2	5.8	17.7	4.9	17.9	11.9	26.5	10.7	24.6
3	6.7	19.0	4.4	17.4	12.4	26.2	8.9	25.3
6	5.8	15.7	4.8	15.4	11.0	24.3	9.8	23.4
10	5.7	14.3	5.3	13.6	10.9	22.8	9.9	21.6

offered very satisfactory levels for all the M's considered. A similar behavior has been observed under ARCH errors.

The empirical levels of one-lag test statistics $Q(l)$ and $S(l)$ are given in Table 2. Under Gaussian errors, the χ^2 distribution of the test statistics $Q(l)$ and $S(l)$ offered a satisfactory approximation for all lags, at the two significance levels, with generally small differences between the two test statistics, for a given l. However, under uncorrelated but dependent errors, the χ^2 approximation appeared inadequate for $S(l)$. Under uncorrelated but one-dependent errors the test statistics $S(l)$ overrejected the null hypothesis severally for the first lag, while $Q(l)$

TAB. 2.2. Empirical levels (in percentage) of the individual lags test statistics $Q(l)$ and $S(l)$ defined by (2.8) and (2.10), for the threshold model given by (2.11) with Gaussian errors, uncorrelated but one-dependent errors and ARCH errors.

	$\alpha = 0.05$				$\alpha = 0.10$			
	$n = 600$		$n = 1200$		$n = 600$		$n = 1200$	
l	$Q(l)$	$S(l)$	$Q(l)$	$S(l)$	$Q(l)$	$S(l)$	$Q(l)$	$S(l)$
	Gaussian errors							
1	6.1	4.9	5.2	5.1	11.9	10.8	10.0	9.8
2	4.8	4.7	5.2	5.0	10.2	9.1	11.3	10.5
3	4.4	4.2	4.6	4.9	9.3	9.7	9.0	8.8
4	5.0	5.1	5.2	5.1	10.0	9.3	11.3	10.6
5	5.6	5.3	5.9	6.2	10.9	10.4	10.9	10.8
	Uncorrelated but one-dependent errors							
1	5.7	12.5	6.0	14.1	11.9	19.3	11.7	22.6
2	5.3	5.5	5.7	5.4	10.6	10.2	11.2	9.9
3	4.5	5.1	4.2	4.2	9.4	10.9	8.2	8.9
4	4.1	4.2	5.1	4.2	9.3	8.6	8.5	9.1
5	4.3	4.4	5.8	5.7	9.3	9.0	10.9	11.4
	ARCH errors							
1	7.3	17.5	4.7	15.2	13.1	25.5	9.4	21.5
2	5.8	12.1	4.8	12.1	11.3	20.1	11.1	19.8
3	4.4	7.1	5.9	8.7	11.1	13.6	10.8	15.2
4	5.8	6.8	4.8	6.6	11.6	12.1	10.6	11.9
5	5.7	5.9	5.4	5.6	10.7	10.7	9.5	10.5

admitted very satisfactory levels for all values of l. Under ARCH errors, $S(l)$ overrejected the null hypothesis for small lags.

Overall, the finite sample performance of the test statistics Q_M and the complementary one-lag test statistics $Q(l)$, $l = 1, \ldots, M$, seemed satisfactory and these test statistics can be recommended for use for diagnosing multivariate nonlinear models with martingale difference errors, which are very common in economic and financial applications. Assuming independent errors and using the portmanteau or one-lag test procedures may lead to incorrect inferences; our empirical study revealed that assuming independent errors when in fact the errors were uncorrelated but dependent may lead to severe distortion in the empirical levels. This highlights the benefits of developing test procedures relying on a weaker assumption for the error term.

5. Conclusion

In this paper we derived the asymptotic distribution of residual autocovariance and autocorrelation matrices for a general class of multivariate nonlinear time series models assuming only that the error term is a martingale difference sequence. We developed two types of applications : global test statistics of the portmanteau type and one-lag test statistics. We observed that the asymptotic distribution of the proposed test statistics may be very different than the one obtained by Li (1992) assuming independent errors. In Monte Carlo experiments, we demonstrated that the proposed portmanteau and one-lag test statistics have reasonable finite sample performances, at least for the models considered. We investigated the classical test statistics for diagnosing threshold time series models. Under independent errors, all the test statistics displayed a satisfactory behavior. However, under uncorrelated but one-dependent errors and under ARCH errors, it appeared that the classical test statistics may be unreliable, overrejecting severely, while the proposed test statistics offered satisfactory levels in most cases. It is hoped that the proposed test statistics studied in this paper will be useful for diagnosing multivariate nonlinear models under martingale difference errors.

Acknowledgements

This research is supported by a grant from the National Science and Engineering Research Council of Canada.

REFERENCES

Chan, K. S. and Tong, H. (1986), 'On estimating thresholds in autoregressive models', *Journal of Time Series Analysis* **7**, 179–190.

Duchesne, P. (2005), 'On the asymptotic distribution of residual autocovariances in VARX models with applications', *Test* **14**, 449–473.

Duchesne, P. (2004), 'On matricial measures of dependence in vector ARCH models with applications to diagnostic checking', *Statistics and Probability Letters* **68**, 149-160.

Francq, C. and Raïssi, H. (2007), 'Multivariate portmanteau test for autoregressive models with uncorrelated but nonindependent errors', *Journal of Time Series Analysis* **28**, 454–470.

Francq, C., Roy, R. and Zakoïan, J. M. (2005), 'Diagnostic checking in ARMA models with uncorrelated errors', *Journal of the American Statistical Association* **100**, 532–544.

Gouriéroux, C. (1997), *ARCH Models and Financial Applications*, Springer : New York.

Granger, C. W. J. and Andersen, A. P. (1978), *An Introduction to Bilinear Time Series Models*, Gottingen : Vandenhack and Ruprecht.

Granger, C. W. J. and Teräsvirta, T. (1993), *Modelling Nonlinear Economic Relationships*, Oxford University Press : Oxford.

Hamilton, J. D. (1989), 'A new approach to the economic analysis of nonstationary time series and the business cycle', *Econometrica* **57**, 357–384.

Harville, D. A. (1997), *Matrix Algebra from a Statistician's Perspective*, Springer-Verlag : New York.

Hosking, J. (1980), 'The multivariate portmanteau statistic', *Journal of the American Statistical Association* **75**, 602–608.

Hwang, S. Y., Basawa, I. V. and Reeves, J. (1994), 'The asymptotic distribution of residual autocorrelations and related tests of fit for a class of nonlinear time series models', *Statistica Sinica* **4**, 107-125.

Klimko, L. A. and Nelson, P. I. (1978), 'On conditional least squares estimation for stochastic processes', *The Annals of Statistics* **6**, 629–642.

Li, W. K. (2004), *Diagnostic Checks in Time Series*, Chapman & Hall/CRC : New York.

Li, W. K. (1992), 'On the asymptotic standard errors of residual autocorrelations in nonlinear time series modelling', *Biometrika* **79**, 435–437.

Li, W. K. and McLeod, A. I. (1981), 'Distribution of the residual autocorrelations in multivariate ARMA time series models', *Journal of the Royal Statistical Society, Series B* **43**, 231–239.

Ling, S. and Li, W. K. (1997), 'Diagnostic checking of nonlinear multivariate time series with multivariate ARCH errors', *Journal of Time Series Analysis* **18**, 447–464.

Ljung, G. M. and Box, G. E. P. (1978), 'On a measure of lack of fit in time series models', *Biometrika* **65**, 297–304.

Romano, J. L. and Thombs, L. A. (1996), 'Inference for autocorrelations under weak assumptions', *Journal of the American Statistical Association* **91**, 590–600.

Serfling, R. J. (1980), *Approximation Theorems of Mathematical Statistics*, Wiley : New York.

Taniguchi, M. and Kakizawa, Y. (2000), *Asymptotic Theory of Statistical Inference for Time Series*, Springer : New York.

Teräsvirta, T. (1994), 'Specification, estimation, and evaluation of smooth transition autoregressive models', *Journal of the American Statistical Association* **89**, 208–218.

Tjøstheim, D. (1986), 'Estimation in nonlinear time series models', *Stochastic Processes and Their Applications* **21**, 251–273.

Tong, H. (1990), *Non-linear Time Series : A Dynamical System Approach*, Oxford University Press : Oxford.

Tong, H. (1983), *Threshold Models in Non-linear Time Series Analysis*, Lecture Notes in Statistics **21**, Springer-Verlag : New York.

Tong, H. (1978), 'On a threshold model', in *Pattern Recognition and Signal Processing*, C. H. Chen (Ed.), Sijtoff and Noordhoff : Amsterdam.

Tong, H. and Lim, K. S. (1980), 'Threshold autoregression, limit cycles and cyclical data' (with discussion), *Journal of the Royal Statistical Society, Series B* **42**, 245–292.

Tsay, R. S. (2005), *Analysis of Financial Time Series*, Second Edition, Wiley : New York.

Tsay, R. S. (1998), 'Testing and modeling multivariate threshold models', *Journal of the American Statistical Association* **93**, 1188–1202.

van Dijk, D., Teräsvirta, T. and Franses, P. H. (2002), 'Smooth transition autoregressive models – A survey of recent developments', *Econometric Reviews* **21**, 1–47.

CONCLUSION

Dans ce mémoire, nous nous sommes intéressés aux modèles non-linéaires en séries chronologiques multivariés. Dans un premier temps, nous avons présenté quelques modèles et nous avons discuté d'une méthode pour estimer ces modèles. Ensuite, nous nous sommes intéressés à l'article de Li (1992), qui obtient la distribution asymptotique des autocorrélations résiduelles pour une classe très générale de modèles non-linéaires univariés. Nous avons généralisé les résultats de Li (1992) dans deux directions importantes : nous avons adapté son résultat au contexte de séries chronologiques multivariées, et ce, en présumant seulement que le terme d'erreur est une différence de martingale, plutôt qu'un bruit blanc fort. À partir de ce résultat théorique, nous avons développé deux statistiques de tests pour vérifier l'adéquation d'un modèle non-linéaire multivarié. Plus précisément, nous avons construit la statistique portemanteau Q_M qui permet de tester la signification conjointe des $\mathbf{\Gamma_a}(l), l = 1, \ldots, M$ et les statistiques à des délais individuels $Q(l)$ qui testent la signification des $\mathbf{\Gamma_a}(l)$ lorsque les erreurs forment une différence de martingale. Finalement, nous avons considéré les statistiques S_M et $S(l)$ obtenues en présumant que les erreurs sont un bruit blanc fort. Pour illustrer nos résultats théoriques, nous avons effectué des études de simulations pour les niveaux en considérant un modèle SETAR multivarié. Ces études visaient à faire ressortir plusieurs éléments, que nous énonçons par ordre d'importance. Premièrement, nous avons constaté que lorsque les erreurs de nos modèles étaient indépendantes, les niveaux empiriques des statistiques de test Q_M, $Q(l)$, S_M et $S(l)$ étaient tous raisonnablement proches des niveaux théoriques, comme attendu. Cependant, lorsque les erreurs étaient dépendantes mais non-corrélées, les niveaux empiriques des statistiques classiques S_M et $S(l)$ affichaient de sévères distorsions, alors que les niveaux empiriques de Q_M et $Q(l)$ continuaient d'être près des niveaux théoriques. Ces conclusions sont sans aucun doute les contributions les plus originales de ce mémoire. Deuxièmement, nous avons constaté que les niveaux empiriques des statistiques corrigées Q_M et $Q(l)$ affichaient de meilleures performances que les statistiques non corrigées H_M et $H(l)$, particulièrement pour les grandes valeurs de M et l, rejoignant ainsi des conclusions obtenues dans la littérature dans d'autres contextes reliés à celui de ce mémoire. Finalement nous avons constaté que les niveaux empiriques se rapprochaient de plus en plus des niveaux théoriques lorsque n augmente. Il est à noter que les tailles échantillonales doivent être assez grandes pour obtenir de bons résultats. Ce constat n'est guère surprenant, étant donné la complexité des modèles, affichant des particularités non-linéaires, dans un contexte de données dépendantes multivariées. Suite à cette recherche, nous espérons que les statistiques de test proposées dans ce mémoire se révèleront des outils grandement utilisés dans la validation de modèles non-linéaires multivariés en séries chronologiques.

Annexe A

DÉMONSTRATIONS DES PRINCIPAUX RÉSULTATS

Dans cette section, quelques démonstrations des principaux résultats obtenus sont présentées. En particulier, la distribution asymptotique des estimateurs des moindres carrés est énoncée sans démonstration dans l'article. Bien que la preuve repose sur des modifications de la preuve originale effectuée dans Tjøstheim (1986) ; voir aussi Taniguchi et Kakizawa (2000, Théorème 3.2.26), par souci de complétude et de rigueur nous la présentons ici en annexe. Le résultat est original et ne semble pas disponible dans la littérature, même si la conclusion est quelque peu anticipée. De plus, la preuve du théorème principal donné dans l'article était présentée plutôt succintement. Nous fournissons ici une preuve détaillée.

A.1. Distribution des estimateurs des paramètres dans un modèle non-linéaire multivarié

Considérons un modèle non-linéaire multivarié de dimension d défini dans l'équation (0.1). Définissons le critère des moindres carrés comme suit :

$$S_n(\boldsymbol{\theta}) = \sum_{t=\tau+1}^{n} \mathbf{a}_t^\top \boldsymbol{\Sigma}_{\mathbf{a}}^{-1} \mathbf{a}_t,$$

ou de manière équivalente :

$$S_n(\boldsymbol{\theta}) = \sum_{t=\tau+1}^{n} \left\{\mathbf{Y}_t - \mathbf{f}(\mathcal{I}_{t-1}, \boldsymbol{\theta})\right\}^\top \boldsymbol{\Sigma}_{\mathbf{a}}^{-1} \left\{\mathbf{Y}_t - \mathbf{f}(\mathcal{I}_{t-1}, \boldsymbol{\theta})\right\}.$$

Rappelons que le choix de la constante τ est discuté dans Taniguchi et Kakizawa (2000, p. 98). Soit $\boldsymbol{\theta}_0$ la vraie valeur de $\boldsymbol{\theta}$. Supposons que $S_n(\boldsymbol{\theta})$ est deux fois différentiable par rapport à $\boldsymbol{\theta}$. Alors, le développement en série de Taylor de $S_n(\boldsymbol{\theta})$ autour de $\boldsymbol{\theta}_0$ est :

$$\begin{aligned} S_n(\boldsymbol{\theta}) &= S_n(\boldsymbol{\theta}_0) + (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^\top \frac{\partial S_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} + \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^\top \frac{\partial^2 S_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) \\ & \quad -\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^\top \left\{ \frac{\partial^2 S_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} - \frac{\partial^2 S_n(\boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \right\} (\boldsymbol{\theta} - \boldsymbol{\theta}_0), \\ &= S_n(\boldsymbol{\theta}_0) + (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^\top \frac{\partial S_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} + \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^\top V_n (\boldsymbol{\theta} - \boldsymbol{\theta}_0) \\ & \quad +\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^\top T_n(\boldsymbol{\theta}^*)(\boldsymbol{\theta} - \boldsymbol{\theta}_0), \end{aligned}$$

où $\boldsymbol{\theta}^*$ est un point intermédiaire entre $\boldsymbol{\theta}$ et $\boldsymbol{\theta}_0$. Le théorème suivant est dû à Klimko et Nelson (1978).

Théorème A.1.1. *Supposons que $\{\mathbf{Y}_t, t \in \mathbb{Z}\}$ et $S_n(\boldsymbol{\theta})$ sont tels que quand $n \longrightarrow \infty$:*

(B1) $n^{-1}\partial S_n(\boldsymbol{\theta}_0)/\partial \boldsymbol{\theta} \stackrel{p.s}{\to} \mathbf{0}$,

(B2) $n^{-1}\mathbf{V}_n \stackrel{p.s}{\to} \mathbf{V}$ *où* $\mathbf{V}$ *est une matrice* $K \times K$ *définie positive, et*

(B3) pour $j, k = 1, \ldots, K$:

$$\lim_{n \to \infty} \sup_{\partial \to 0} (n\partial)^{-1} |T_n(\boldsymbol{\theta}^*)_{jk}| < \infty \ p.s,$$

où, $T_n(\boldsymbol{\theta}^*)_{jk}$ *est la* (j,k) *ième composante de* $T_n(\boldsymbol{\theta}^*)$.

Alors, il existe une suite d'estimateurs $\hat{\boldsymbol{\theta}}_n$ *tel que* $\hat{\boldsymbol{\theta}}_n \stackrel{p.s}{\to} \boldsymbol{\theta}_0$, *et pour tout* $\varepsilon > 0$, *il existe un évènement* E *avec* $P(E) > 1 - \varepsilon$ *et un* n_0 *tel que sur* E, *pour* $n > n_0$, $\partial S_n(\hat{\boldsymbol{\theta}}_n)/\partial \boldsymbol{\theta} = \mathbf{0}$ *et* $S_n(\boldsymbol{\theta})$ *atteint un minimum relatif à* $\hat{\boldsymbol{\theta}}_n$. *De plus si :*

(B4) $n^{-1/2}\partial S_n(\boldsymbol{\theta}^0)/\partial \boldsymbol{\theta} \stackrel{d}{\to} N(\mathbf{0}, \mathbf{W})$, *pour une certaine matrice de variance* $\mathbf{W}$, *alors :*

$$n^{1/2}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \stackrel{d}{\to} N(\mathbf{0}, \mathbf{V}^{-1}\mathbf{W}\mathbf{V}^{-1}).$$

On peut maintenant trouver la distribution asymptotique de $\hat{\boldsymbol{\theta}}_n$. Pour faciliter la présentation de la preuve, nous énonçons à nouveau le théorème que nous voulons démontrer.

Théorème A.1.2. *Soit $\{\mathbf{Y}_t, t \in \mathbb{Z}\}$ un processus strictement stationnaire et ergodique tel que $E\{||\mathbf{Y}_t||^2\} < \infty$ et où $\mathbf{f}(\mathcal{I}_{t-1}, \boldsymbol{\theta})$ est trois fois différentiable p.s. sur un ensemble ouvert B contenant la vraie valeur $\boldsymbol{\theta}_0$. De plus, on suppose que les conditions suivantes sont respectées :*

(C1) Pour $j, l = 1, \ldots, K$:

$$E\left[\left|\left|\frac{\partial}{\partial\theta_j}\left\{\mathbf{f}(\mathcal{I}_{t-1}, \boldsymbol{\theta}_0)^\top \boldsymbol{\Sigma}_{\mathbf{a}}^{-1}\mathbf{f}(\mathcal{I}_{t-1}, \boldsymbol{\theta}_0)\right\}\right|\right|\right] < \infty,$$

et

$$E\left[\left|\left|\frac{\partial^2}{\partial\theta_j\partial\theta_l}\left\{\mathbf{f}(\mathcal{I}_{t-1}, \boldsymbol{\theta}_0)^\top \boldsymbol{\Sigma}_{\mathbf{a}}^{-1}\mathbf{f}(\mathcal{I}_{t-1}, \boldsymbol{\theta}_0)\right\}\right|\right|\right] < \infty.$$

(C2) Pour $j = 1, \ldots, K$, les vecteurs

$$\boldsymbol{\Sigma}_{\mathbf{a}}^{-1/2}\frac{\partial \mathbf{f}(\mathcal{I}_{t-1}, \boldsymbol{\theta}_0)}{\partial\theta_j},$$

sont linéairement indépendants au sens de l'algèbre linéaire.

(C3) Pour $\boldsymbol{\theta} \in B$, il existe une fonction $H_t^{ijl}(\mathbf{Y}_1, ..., \mathbf{Y}_{t-1})$ telle que :

$$\left|\frac{\partial^3}{\partial\theta_i\partial\theta_j\partial\theta_l}\left\{\mathbf{f}(\mathcal{I}_{t-1}, \boldsymbol{\theta})^\top \boldsymbol{\Sigma}_{\mathbf{a}}^{-1}\mathbf{f}(\mathcal{I}_{t-1}, \boldsymbol{\theta})\right\}\right| \leq H_t^{ijl},$$

avec

$$E(H_t^{ijl}) < \infty,$$

pour $i, j, l = 1, \ldots, K$.

(C4) De plus, la matrice suivante existe :

$$\mathbf{R} = E\left[\left\{\frac{\partial \mathbf{f}(\mathcal{I}_{t-1}, \boldsymbol{\theta}_0)^\top}{\partial\boldsymbol{\theta}}\boldsymbol{\Sigma}_{\mathbf{a}}^{-1}\mathbf{a}_t\right\}\left\{\mathbf{a}_t^\top\boldsymbol{\Sigma}_{\mathbf{a}}^{-1}\frac{\partial \mathbf{f}(\mathcal{I}_{t-1}, \boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}^\top}\right\}\right] < \infty.$$

Alors il existe une suite d'estimateurs $\hat{\boldsymbol{\theta}}_n$ telle que $\hat{\boldsymbol{\theta}}_n \xrightarrow{p.s} \boldsymbol{\theta}_0$, et pour tout $\varepsilon > 0$, il existe un évènement E avec $P(E) > 1 - \varepsilon$ et un n_0 tel que sur E, pour $n > n_0$, $\partial S_n(\hat{\boldsymbol{\theta}}_n)/\partial\boldsymbol{\theta} = 0$ et $S_n(\boldsymbol{\theta})$ atteint un minimum relatif à $\hat{\boldsymbol{\theta}}_n$. De plus,

$$n^{1/2}\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right) \xrightarrow{d} N(\mathbf{0}, \mathbf{U}^{-1}\mathbf{R}\mathbf{U}^{-1}),$$

où

$$\mathbf{U} = E\left\{\frac{\partial \mathbf{f}(\mathcal{I}_{t-1}, \boldsymbol{\theta}_0)^\top}{\partial\boldsymbol{\theta}}\boldsymbol{\Sigma}_{\mathbf{a}}^{-1}\frac{\partial \mathbf{f}(\mathcal{I}_{t-1}, \boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}^\top}\right\},$$

et $\mathbf{R}$ *est défini plus haut.*

On note que si $\{\mathbf{a}_t\}$ est un bruit blanc fort, alors :

$$n^{1/2}\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right) \xrightarrow{d} N(\mathbf{0}, \mathbf{U}^{-1}). \tag{A.1.1}$$

Preuve. Commençons par calculer la dérivée de notre critère. La dérivée première est :

$$\begin{aligned}\frac{\partial S_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}^\top} &= 2\sum_{t=1}^{n} \mathbf{a}_t^\top \boldsymbol{\Sigma}_{\mathbf{a}}^{-1} \frac{\partial \mathbf{a}_t}{\partial \boldsymbol{\theta}}, \\ &= -2\sum_{t=1}^{n} \mathbf{a}_t^\top \boldsymbol{\Sigma}_{\mathbf{a}}^{-1} \frac{\partial \mathbf{f}(\mathcal{I}_{t-1}, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}. \end{aligned} \tag{A.1.2}$$

De l'hypothèse que le processus soit strictement stationnaire et ergodique, du Théorème 1.2.2, de l'équation (A.1.2) et de l'hypothèse *(C1)* on obtient :

$$n^{-1}\frac{\partial S_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \overset{p.s}{\rightarrow} -2E\left[\sum_{t=1}^{n} \frac{\partial \mathbf{f}(\mathcal{I}_{t-1}, \boldsymbol{\theta}_0)^\top}{\partial \boldsymbol{\theta}} \boldsymbol{\Sigma}_{\mathbf{a}}^{-1} \left\{\mathbf{Y}_t - \mathbf{f}(\mathcal{I}_{t-1}, \boldsymbol{\theta}_0)\right\}\right] = \mathbf{0}.$$

Alors, l'hypothèse *(B1)* du Théorème A.1.1 est satisfaite. Pour la dérivée seconde du critère on a :

$$\begin{aligned}\frac{\partial^2 S_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^\top} &= 2\sum_{t=1}^{n} \frac{\partial \mathbf{a}_t}{\partial \boldsymbol{\theta}} \boldsymbol{\Sigma}_{\mathbf{a}}^{-1} \frac{\partial \mathbf{a}_t}{\partial \boldsymbol{\theta}^\top} + 2\sum_{t=1}^{n} \left(\mathbf{a}_t^\top \boldsymbol{\Sigma}_{\mathbf{a}}^{-1} \otimes \mathbf{I}_K\right) \frac{\partial}{\partial \boldsymbol{\theta}^\top} \mathrm{vec}\left(\frac{\partial \mathbf{a}_t^\top}{\partial \boldsymbol{\theta}}\right), \\ &= 2\sum_{t=1}^{n} \frac{\partial \mathbf{f}(\mathcal{I}_{t-1}, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \boldsymbol{\Sigma}_{\mathbf{a}}^{-1} \frac{\partial \mathbf{f}(\mathcal{I}_{t-1}, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}^\top} \\ &\quad +2\sum_{t=1}^{n} \left(\mathbf{a}_t^\top \boldsymbol{\Sigma}_{\mathbf{a}}^{-1} \otimes \mathbf{I}_K\right) \frac{\partial}{\partial \boldsymbol{\theta}^\top} \mathrm{vec}\left(-\frac{\partial \mathbf{f}(\mathcal{I}_{t-1}, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}^\top}\right).\end{aligned}$$

Pour deux matrices $\mathbf{A}$ et $\mathbf{B}$, $\mathbf{A} \otimes \mathbf{B}$ dénote le produit de Kronecker entre ces deux matrices. Harville (1997) discute des propriétés entourant le produit de Kronecker. Or, en invoquant le théorème de la double espérance on remarque que :

$$\begin{aligned} E\left\{2\sum_{t=1}^{n} \mathbf{a}_t^\top \boldsymbol{\Sigma}_{\mathbf{a}}^{-1} \otimes \mathbf{I}_K \frac{\partial}{\partial \boldsymbol{\theta}^\top} \mathrm{vec}\left(\frac{\partial \mathbf{a}_t^\top}{\partial \boldsymbol{\theta}}\right)\right\} &= \\ E\left[E\left\{2\sum_{t=1}^{n} \left(\mathbf{a}_t^\top \boldsymbol{\Sigma}_{\mathbf{a}}^{-1} \otimes \mathbf{I}_K\right) \frac{\partial}{\partial \boldsymbol{\theta}^\top} \mathrm{vec}\left(\frac{\partial \mathbf{a}_t^\top}{\partial \boldsymbol{\theta}}\right) | \mathcal{I}_{t-1}\right\}\right] &= \mathbf{0}.\end{aligned}$$

En utilisant ce résultat, on obtient de l'hypothèse que le processus soit strictement stationnaire et ergodique, du Théorème 1.2.2 et de l'hypothèse *(B1)* que :

$$n^{-1}\frac{\partial^2 S_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} \xrightarrow{p.s} 2\mathbf{U}.$$

En utilisant le fait que les vecteurs $\boldsymbol{\Sigma}_{\mathbf{a}}^{-1/2}\partial\mathbf{f}(\mathcal{I}_{t-1},\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}_j$, $j = 1, ..., K$, sont linéairement indépendants (hypothèse *(C2)*), on peut montrer que $\mathbf{U}$ est définie positive, ce qui vérifie l'hypothèse *(B2)*. De même, l'hypothèse *(B3)* est vérifiée à partir de l'hypothèse *(C3)* et du Théorème 1.2.2. Finalement, l'hypothèse *(B4)* du Théorème A.1.1 découle directement du Théorème 1.2.3, ce qui conclut la preuve. La démonstration du résultat énoncé dans l'équation (A.1.1) découle du fait que si $\{\mathbf{a}_t\}$ est un bruit blanc fort, alors $\mathbf{R} = \mathbf{U}$.

A.2. Distribution des autocovariances résiduelles dans un modèle non-linéaire multivarié

Commençons par trouver la distribution asymptotique de $\mathbf{c}_{\mathbf{a}}$.

Lemme A.2.1. *Si* $\{\mathbf{a}_t, t \in \mathbb{Z}\}$ *est une différence de martingale de moyenne zéro, alors :*

$$n^{1/2}\mathbf{c}_{\mathbf{a}} \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Delta}_M),$$

où

$$\boldsymbol{\Delta}_M = (\boldsymbol{\Delta}_{ij})\,,\ \ i,j = 1,\ldots,M,$$

avec

$$\boldsymbol{\Delta}_{ij} = E(\mathbf{a}_{t-i}\mathbf{a}_{t-j}^\top \otimes \mathbf{a}_t\mathbf{a}_t^\top), i,j = 1,\ldots,M.$$

Lorsque $\{\mathbf{a}_t\}$ est une suite de vecteurs aléatoires indépendants, la matrice de variance asymptotique se simplifie de la manière suivante :

$$\boldsymbol{\Delta}_M = \mathbf{I}_M \otimes \boldsymbol{\Sigma}_{\mathbf{a}} \otimes \boldsymbol{\Sigma}_{\mathbf{a}}.$$

Le résultat présumant que $\{\mathbf{a}_t\}$ est composé de vecteurs aléatoires indépendants est démontré dans Li et McLeod (1981). Pour démontrer notre théorème principal, nous avons également besoin de trois lemmes.

Lemme A.2.2.

$$\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 = -\frac{1}{2}\left\{\sum_{t=1}^{n} \frac{\partial \mathbf{f}(\mathcal{I}_{t-1},\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}^\top \boldsymbol{\Sigma}_{\mathbf{a}}^{-1} \frac{\partial \mathbf{f}(\mathcal{I}_{t-1},\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\right\}^{-1} \mathbf{S}_n^c + \mathbf{o}_P(n^{-1}),$$

où $\mathbf{S}_n^c = -2\sum_{t=1}^n \partial \mathbf{f}(\mathcal{I}_{t-1},\boldsymbol{\theta})/\partial\boldsymbol{\theta}^\top \boldsymbol{\Sigma}_{\mathbf{a}}^{-1}\mathbf{a}_t$.

Preuve. Considérons un développement en série de Taylor de $n^{-1/2}\partial S_n(\hat{\boldsymbol{\theta}}_n)/\partial\boldsymbol{\theta}$ autour de $\boldsymbol{\theta}_0$, on obtient alors :

$$\begin{aligned}
\mathbf{0} &= n^{-1/2}\frac{\partial S_n(\hat{\boldsymbol{\theta}}_n)}{\partial \boldsymbol{\theta}}, \\
&= n^{-1/2}\frac{\partial S_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} + \frac{1}{n}\left\{\frac{\partial^2 S_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} + T_n(\boldsymbol{\theta}^*)\right\} n^{1/2}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0),
\end{aligned}$$

où $\boldsymbol{\theta}^*$ est un point intermédiaire entre $\boldsymbol{\theta}$ et $\boldsymbol{\theta}_0$. Comme $n^{-1/2}T_n(\boldsymbol{\theta}^*)(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ est $\mathbf{o}_P(1)$, on a :

$$\begin{aligned}
\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 &= \left\{\frac{\partial^2 S_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\right\}^{-1} \frac{\partial S_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} + \mathbf{o}_P(n^{-1/2}), \\
&= \left\{\sum_{t=1}^{n} \frac{\partial \mathbf{f}(\mathcal{I}_{t-1},\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}^\top \boldsymbol{\Sigma}_{\mathbf{a}}^{-1} \frac{\partial \mathbf{f}(\mathcal{I}_{t-1},\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\right\}^{-1} \sum_{t=1}^{n} \frac{\partial \mathbf{f}(\mathcal{I}_{t-1},\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}^\top \boldsymbol{\Sigma}_{\mathbf{a}}^{-1}\mathbf{a}_t \\
&\quad + \mathbf{o}_P(n^{-1/2}).
\end{aligned}$$

Utilisant la définition de $\mathbf{S}_n^c$ nous obtenons le résultat.

Lemme A.2.3. $\mathbf{c}_{\hat{\mathbf{a}}} = \mathbf{c}_{\mathbf{a}} + \mathbf{J}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + \mathbf{o}_P(n^{-1/2})$, *où*

$$\mathbf{J}^\top = -\frac{1}{n}E\left(\mathbf{J}_1^\top, \ldots, \mathbf{J}_M^\top\right)^\top,$$

avec

$$\mathbf{J}_k = \sum_{t=1+k}^{n} (\mathbf{a}_{t-k} \otimes \mathbf{I}_d)\frac{\partial \mathbf{f}(\mathcal{I}_{t-1},\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}^\top}.$$

Preuve. Considérons un développement en série de Taylor de $\mathbf{c}_{\mathbf{a}}$ autour de $\boldsymbol{\theta}_0$ évalué en $\hat{\boldsymbol{\theta}}_n$:

$$\mathbf{c}_{\hat{\mathbf{a}}} = \mathbf{c}_{\mathbf{a}} + \frac{\partial \mathbf{c}_{\mathbf{a}}}{\partial\boldsymbol{\theta}^\top}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + \mathbf{o}_P(n^{-1/2}).$$

Il reste à expliciter la matrice de dérivées de $\partial\mathbf{c}_{\mathbf{a}}(i)/\partial\boldsymbol{\theta}^\top$. Notons que :

$$\mathbf{c}_{\mathbf{a}}(i) = \frac{1}{n}\sum_{t=i+1}^{n} \mathrm{vec}(\mathbf{a}_t\mathbf{a}_{t-i}^\top).$$

Ainsi en dérivant par rapport à $\boldsymbol{\theta}$:

$$\begin{aligned}\frac{\partial \mathbf{c}_{\mathbf{a}}(i)}{\partial \boldsymbol{\theta}^{\top}} &= \frac{1}{n}\sum_{t=i+1}^{n}\frac{\partial \mathrm{vec}(\mathbf{a}_t\mathbf{a}_{t-i}^{\top})}{\partial \boldsymbol{\theta}^{\top}},\\ &= \frac{1}{n}\sum_{t=i+1}^{n}\left\{(\mathbf{I}_d\otimes \mathbf{a}_t)\frac{\partial \mathbf{a}_{t-i}}{\partial \boldsymbol{\theta}^{\top}}+(\mathbf{a}_{t-i}\otimes \mathbf{I}_d)\frac{\partial \mathbf{a}_t}{\partial \boldsymbol{\theta}^{\top}}\right\}.\end{aligned}$$

Or, en invoquant de nouveau le théorème de la double espérance on remarque que :

$$\begin{aligned}E\left\{\frac{1}{n}\sum_{t=i+1}^{n}(\mathbf{I}_d\otimes \mathbf{a}_t)\frac{\partial \mathbf{a}_{t-i}}{\partial \boldsymbol{\theta}^{\top}}\mathbf{a}_{t-i}\right\} &=\\ E\left[E\left\{\frac{1}{n}\sum_{t=i+1}^{n}(\mathbf{I}_d\otimes \mathbf{a}_t)\frac{\partial \mathbf{a}_{t-i}}{\partial \boldsymbol{\theta}^{\top}}|\mathcal{I}_{t-1}\right\}\right] &= \mathbf{0}.\end{aligned}$$

On a donc que :

$$\frac{\partial \mathbf{c}_{\mathbf{a}}(i)}{\partial \boldsymbol{\theta}^{\top}} \xrightarrow{p} -E\left\{(\mathbf{a}_{t-i}\otimes \mathbf{I}_d)\frac{\partial \mathbf{f}(\mathcal{I}_{t-1},\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}^{\top}}\right\}.$$

Des manipulations simples permettent de conclure que :

$$\mathbf{c}_{\hat{\mathbf{a}}} = \mathbf{c}_{\mathbf{a}} + \frac{\partial \mathbf{c}_{\mathbf{a}}}{\partial \boldsymbol{\theta}^{\top}}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + \mathbf{o}_P(n^{-1/2}),$$

étant donné que $\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 = \mathbf{o}_P(n^{-1/2})$ et que $\partial \mathbf{c}_{\mathbf{a}}(i)/\partial \boldsymbol{\theta}^{\top}$ moins son espérance converge vers $\mathbf{0}$ à la vitesse $n^{-1/2}$.

Lemme A.2.4. *La covariance asymptotique entre $n^{1/2}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ et $n^{1/2}\mathbf{c}_{\mathbf{a}}$ est égale à $\mathbf{U}^{-1}\mathbf{J}^{*\top}$, où :*

$$\mathbf{J}^{*\top} = -\frac{1}{n}E\left(\mathbf{J}_1^{*\top},\ldots,\mathbf{J}_M^{*\top}\right)^{\top},$$

avec

$$\mathbf{J}_k^{*\top} = \sum_{t=1+k}^{n}(\mathbf{a}_{t-k}^{\top}\otimes \mathbf{I}_d)\frac{\partial \mathbf{f}(\mathcal{I}_{t-1},\boldsymbol{\theta}_0)^{\top}}{\partial \boldsymbol{\theta}}\boldsymbol{\Sigma}_{\mathbf{a}}^{-1}\mathbf{a}_t\mathbf{a}_t^{\top}.$$

Preuve. Pour démontrer le résultat, il suffit de calculer la covariance pour un des délais.

$$\begin{aligned}
nE\left\{\mathbf{S}_n^c \mathbf{c}_{\mathbf{a}}^\top(i)\right\} &= nE\left\{-2\sum_{t=1}^{n} \frac{\partial \mathbf{f}(\mathcal{I}_{t-1},\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}^\top}\boldsymbol{\Sigma}_{\mathbf{a}}^{-1}\mathbf{a}_t \frac{1}{n}\left(\sum_{j=i+1}^{n}\mathrm{vec}(\mathbf{a}_t\mathbf{a}_{t-j}^\top)\right)^\top\right\}, \\
&= -2E\left\{\sum_{t=2}^{n}\left(\frac{\partial \mathbf{f}(\mathcal{I}_{t-1},\boldsymbol{\theta}_0)^\top}{\partial \boldsymbol{\theta}}\boldsymbol{\Sigma}_{\mathbf{a}}^{-1}\mathbf{a}_t\right)\left(\mathbf{a}_{t-i}^\top \otimes \mathbf{a}_t^\top\right)\right\}, \\
&= -2E\left\{\sum_{t=2}^{n}(\mathbf{a}_{t-i}^\top \otimes \mathbf{I}_d)\frac{\partial \mathbf{f}(\mathcal{I}_{t-1},\boldsymbol{\theta}_0)^\top}{\partial \boldsymbol{\theta}}\boldsymbol{\Sigma}_{\mathbf{a}}^{-1}\mathbf{a}_t\mathbf{a}_t^\top\right\}.
\end{aligned}$$

En utilisant le lemme A.2.2 on obtient bien que :

$$nE\left\{(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\mathbf{c}_{\mathbf{a}}^\top(i)\right\} = \mathbf{U}^{-1}E\left\{\sum_{t=2}^{n}(\mathbf{a}_{t-i}^\top \otimes \mathbf{I}_d)\frac{\partial \mathbf{f}(\mathcal{I}_{t-1},\boldsymbol{\theta}_0)^\top}{\partial \boldsymbol{\theta}}\boldsymbol{\Sigma}_{\mathbf{a}}^{-1}\mathbf{a}_t\mathbf{a}_t^\top\right\}.$$

Ceci conclut la preuve. Nous pouvons maintenant énoncer notre théorème principal.

Théorème A.2.1. *Soit une série chronologique définie par l'équation (0.1). Alors,*

$$n^{1/2}\mathbf{c}_{\hat{\mathbf{a}}} \xrightarrow{d} N(\mathbf{0},\boldsymbol{\Omega}),$$

où $\boldsymbol{\Omega} = \boldsymbol{\Delta}_M - \mathbf{J}\mathbf{U}^{-1}\mathbf{J}^{*\top} - \mathbf{J}^*\mathbf{U}^{-1}\mathbf{J}^\top + \mathbf{J}\mathbf{U}^{-1}\mathbf{R}\mathbf{U}^{-1}\mathbf{J}^\top$ *avec* $\mathbf{J}$, $\mathbf{U}^{-1}$ *et* $\mathbf{J}^*$ *définies respectivement dans le lemme A.2.3, dans le théorème A.1.2 et dans le lemme A.2.4.*

Preuve. En utilisant le lemme A.2.3 on obtient que $n\mathrm{Var}(\hat{\mathbf{c}}_{\mathbf{a}})$ et $n\mathrm{Var}\left\{\mathbf{c}_{\mathbf{a}} - \mathbf{J}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\right\}$ ont la même limite et :

$$\begin{aligned}
n\mathrm{Var}\left\{\mathbf{c}_{\mathbf{a}} - \mathbf{J}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\right\} &= nE\left[\left\{\mathbf{c}_{\mathbf{a}} - \mathbf{J}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\right\}\left\{\mathbf{c}_{\mathbf{a}} - \mathbf{J}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\right\}^\top\right], \\
&= nE\left[(\mathbf{c}_{\mathbf{a}}\mathbf{c}_{\mathbf{a}}^\top) - \mathbf{J}\left\{(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\mathbf{c}_{\mathbf{a}}^\top\right\}\right] \\
&\quad + nE\left[-\left\{\mathbf{c}_{\mathbf{a}}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\right\}\mathbf{J}^\top + \mathbf{J}\mathrm{Var}(\hat{\boldsymbol{\theta}}_n)\mathbf{J}^\top\right], \\
&= \boldsymbol{\Delta}_M - \mathbf{J}\mathbf{U}^{-1}\mathbf{J}^{*\top} - \mathbf{J}^*\mathbf{U}^{-1}\mathbf{J}^\top + \mathbf{J}\mathbf{U}^{-1}\mathbf{R}\mathbf{U}^{-1}\mathbf{J}^\top.
\end{aligned}$$

La dernière égalité découle du lemme A.2.4 et du lemme A.2.2.

Annexe B

SIMULATIONS COMPLÉMENTAIRES

Dans ce chapitre, nous avons effectué des simulations supplémentaires pour mettre en évidence la différence entre les statistiques corrigées pour les niveaux Q_M et $Q(l)$ et les statistiques non corrigées pour les niveaux H_M et $H(l)$, définies respectivement dans les équations (1.5.4), (1.5.2), (1.5.3) et (1.5.1). Plusieurs auteurs ont traitées des différences entre ces statistiques. Entre autres, Duchesne (2005) mentionne que les niveaux empiriques des statistiques corrigées sont meilleurs que pour les statistiques non corrigées, en particulier pour de grandes valeurs de M et l. Pour cette étude, nous avons considéré le même modèle que dans l'article avec différentes valeurs pour les paramètres. Plus précisement, nous avons utilisé le modèle suivant :

$$\mathbf{Y}_t = \begin{cases} \mathbf{\Phi}_1^{(1)}\mathbf{Y}_{t-1} + \mathbf{a}_t, & Y_{t-1}(1) < 0, \\ \mathbf{\Phi}_1^{(2)}\mathbf{Y}_{t-1} + \mathbf{a}_t, & Y_{t-1}(1) \geq 0, \end{cases} \tag{B.1}$$

où

$$\mathbf{\Phi}_1^{(1)} = \begin{pmatrix} 0.5 & 0.2 \\ 0.2 & 0.5 \end{pmatrix}, \ \mathbf{\Phi}_1^{(2)} = \begin{pmatrix} -0.5 & -0.2 \\ -0.2 & -0.5 \end{pmatrix}.$$

Pour le terme d'erreur, nous avons considéré les trois mêmes processus que dans l'article en changeant de nouveau les valeurs des paramètres. Premièrement, nous avons inclus un terme d'erreur Gaussien $\mathbf{a}_4 = \{\mathbf{a}_t\}$, tel que $\mathbf{a}_t \sim N_2(\mathbf{0}, \mathbf{\Sigma_a})$, avec :

$$\mathbf{\Sigma_a} = \begin{pmatrix} 1.0 & -0.3 \\ -0.3 & 1.0 \end{pmatrix}.$$

Ensuite, nous avons considéré $\mathbf{a}_5 = \{\mathbf{a}_t\}$ avec $\mathbf{a}_t = \boldsymbol{\eta}_t\eta_{t-1}(2)$, où $\boldsymbol{\eta} = \{\boldsymbol{\eta}_t, t \in \mathbb{Z}\}$, $\boldsymbol{\eta}_t = (\eta_t(1), \eta_t(2))^\top$, est un bruit blanc Gaussien tel que : $\boldsymbol{\eta}_t \sim N_2(\mathbf{0}, \mathbf{I}_2)$.

Finalement, nous avons inclus deux termes d'erreurs ARCH, soit un modèle à structure de variance conditionnelle diagonale $\mathbf{a}_6 = \{\mathbf{a}_t\}$, où $\mathbf{a}_t = \mathbf{H}_t\boldsymbol{\eta}_t$, $\mathbf{H}_t = \operatorname{diag}(H_t(1), H_t(2))$, $H_t(1) = \{0.01 + 0.03a_{t-1}^2(1)\}^{1/2}$, $H_t(2) = \{0.01 + 0.02a_{t-1}^2(2)\}^{1/2}$, et un modèle à structure de variance conditionnelle non-diagonale $\mathbf{a}_7 = \{\mathbf{a}_t\}$, où $\mathbf{a}_t = \mathbf{H}_t\boldsymbol{\eta}_t$, avec $\mathbf{H}_t$ une matrice deux par deux satisfaisant $\mathbf{V}_t = \mathbf{H}_t\mathbf{H}_t^\top$ avec $V_t(1,1) = 0.01 + 0.3a_{t-1}^2(1)$, $V_t(2,2) = 0.005 + 0.4a_{t-1}^2(2) + 0.06a_{t-2}^2(2)$, $V_t(1,2) = V_t(2,1) = 0.2a_{t-1}(1)a_{t-1}(2)$.

Nous avons étudié les niveaux empiriques des statistiques Q_M, H_M, $Q(l)$ et $H(l)$ aux niveaux nominaux 5% et 10% pour deux tailles échantillonales différentes ($n = 600, 1200$). Pour chaque série, 1000 réalisations indépendantes ont été générées. Les statistiques portemanteaux Q_M et H_M ont été calculées pour les valeurs $M = 1, 2, 3, 6, 10, 15$ et 20. Les statistiques à délais individuels $Q(l)$ et $H(l)$ ont été calculées pour les valeurs $l = 1, \ldots, 8$.

En analysant les résultats présentés dans les tableaux B.1, B.2, B.3 et B.4, on remarque que les statistiques corrigées pour les niveaux Q_M et $Q(l)$ affichent de meilleurs niveaux empiriques que les statistiques non corrigées pour les niveaux H_M et $H(l)$. Ce comportement est particulièrement évident lorsque l'on compare Q_M et H_M, pour $M = 10, 15$ et 20.

TAB. B.1. Niveaux empiriques (en pourcentage) des statistiques de test portemanteau Q_M et H_M défini par (1.5.4) et (1.5.3), pour le modèle (B.1) avec erreurs gaussiennes, erreurs non-corrélées mais dépendantes et erreurs ARCH.

	$\alpha = 0.05$				$\alpha = 0.10$			
	$n = 600$		$n = 1200$		$n = 600$		$n = 1200$	
M	Q_M	H_M	Q_M	H_M	Q_M	H_M	Q_M	H_M
	Erreurs gaussiennes							
1	5.7	5.6	5.5	5.2	11.0	10.7	10.4	10.3
2	4.6	4.6	5.3	5.0	9.9	9.6	10.4	10.1
3	5.1	5.0	5.9	5.6	10.5	10.2	10.8	10.5
6	4.4	4.2	5.5	5.2	9.1	9.0	10.4	10.4
10	4.8	4.8	5.0	4.8	10.0	9.9	9.8	9.6
15	5.0	4.8	5.1	4.7	10.3	9.8	10.1	9.6
20	4.7	4.3	4.9	4.4	9.6	9.2	9.5	9.0
	Erreurs non-corrélées mais dépendantes							
1	6.3	6.1	5.7	5.5	11.8	11.5	11.1	10.9
2	5.8	5.7	5.5	5.5	11.2	11.1	11.0	10.9
3	5.4	5.4	5.2	5.1	10.5	10.5	10.4	10.2
6	4.7	4.6	5.1	5.0	9.5	9.4	10.4	10.2
10	4.8	4.6	4.8	4.7	9.8	9.6	10.0	9.7
15	4.5	4.3	5.1	4.8	9.4	9.2	9.8	9.7
20	4.9	4.7	4.7	4.6	9.8	9.7	9.5	9.4

TAB. B.2. Niveaux empiriques (en pourcentage) des statistiques de test portemanteau Q_M et H_M défini par (1.5.4) et (1.5.3), pour le modèle (B.1) avec erreurs ARCH à structure de variance conditionnelle diagonale (ARCH-D) et non-diagonale (ARCH-ND).

	$\alpha = 0.05$				$\alpha = 0.10$			
	$n = 600$		$n = 1200$		$n = 600$		$n = 1200$	
M	Q_M	H_M	Q_M	H_M	Q_M	H_M	Q_M	H_M
	Erreurs ARCH-D							
1	7.2	7.1	5.6	5.5	12.9	12.7	11.0	10.9
2	6.2	6.1	5.2	5.0	11.5	11.2	10.3	10.1
3	5.8	5.7	4.7	4.6	11.0	10.8	9.7	9.6
6	5.4	5.3	5.1	4.9	10.7	10.7	9.8	9.8
10	5.1	5.0	5.3	5.2	10.3	10.1	10.4	10.3
15	4.9	4.7	5.0	4.8	10.2	10.0	10.1	9.8
20	4.8	4.7	5.1	4.8	9.8	9.6	9.9	9.8
	Erreurs ARCH-ND							
1	5.6	5.4	5.5	5.5	10.3	10.0	10.8	10.8
2	5.5	5.5	5.2	5.0	10.5	10.2	10.0	9.9
3	5.4	5.3	4.7	4.6	10.2	10.0	10.2	10.1
6	4.0	3.9	5.1	4.9	9.4	9.1	9.7	9.4
10	4.0	3.8	5.3	5.2	7.3	6.8	9.7	9.1

TAB. B.3. Niveaux empiriques (en pourcentage) des statistiques de test à délais individuels $Q(l)$ et $H(l)$ défini par (1.5.2) et (1.5.1), pour le modèle (B.1) avec erreurs gaussiennes et erreurs non-corrélées mais dépendantes

	$\alpha = 0.05$				$\alpha = 0.10$			
	$n = 600$		$n = 1200$		$n = 600$		$n = 1200$	
l	$Q(l)$	$H(l)$	$Q(l)$	$H(l)$	$Q(l)$	$H(l)$	$Q(l)$	$H(l)$
	Erreurs gaussiennes							
1	5.7	5.6	5.5	5.2	11.0	10.7	10.4	10.3
2	4.2	4.0	4.9	4.9	9.0	8.8	10.2	10.0
3	5.4	5.2	5.6	5.6	10.3	10.2	10.9	10.8
4	4.3	4.1	4.8	4.8	8.9	8.6	9.7	9.5
5	4.5	4.4	5.1	5.0	9.4	9.4	10.2	10.1
6	4.8	4.8	5.5	5.5	10.0	9.8	10.5	10.4
7	5.2	5.0	4.9	4.7	10.1	9.7	10.0	9.6
8	4.8	4.6	5.3	5.2	9.7	9.6	10.3	10.2
	Erreurs non-corrélées mais dépendantes							
1	6.3	6.1	5.7	5.5	11.8	11.5	11.1	10.9
2	5.4	5.4	5.5	5.4	10.6	10.5	10.5	10.5
3	4.9	4.8	5.1	5.0	9.6	9.6	9.9	9.8
4	5.5	5.3	4.4	4.3	10.8	10.5	9.2	9.1
5	4.7	4.7	4.9	4.7	9.7	9.5	10.0	9.8
6	4.2	4.1	5.2	5.0	9.0	8.9	10.3	10.1
7	4.8	4.6	4.5	4.4	9.9	9.8	9.6	9.6
8	5.1	4.9	4.8	4.8	10.1	9.8	9.9	9.8

TAB. B.4. Niveaux empiriques (en pourcentage) des statistiques de test à délais individuels $Q(l)$ et $H(l)$ défini par (1.5.2) et (1.5.1), pour le modèle (B.1) avec erreurs ARCH à structure de variance conditionnelle diagonale (ARCH-D) et non-diagonale (ARCH-ND).

	$\alpha = 0.05$				$\alpha = 0.10$			
	$n = 600$		$n = 1200$		$n = 600$		$n = 1200$	
l	$Q(l)$	$H(l)$	$Q(l)$	$H(l)$	$Q(l)$	$H(l)$	$Q(l)$	$H(l)$
	Erreurs ARCH-D							
1	7.2	7.1	5.6	5.5	13.1	12.9	11.3	11.2
2	5.5	5.5	4.9	4.9	11.0	10.9	10.0	9.9
3	5.8	5.6	4.4	4.3	11.1	11.0	9.2	9.0
4	4.6	4.5	5.1	5.0	9.3	9.2	10.3	10.3
5	5.1	5.1	5.3	5.1	10.3	10.3	10.5	10.3
6	4.7	4.5	4.8	4.7	9.6	9.2	9.7	9.5
7	5.3	5.0	4.3	4.1	10.1	10.0	9.4	9.3
8	4.8	4.6	4.9	4.8	9.5	9.4	9.6	9.6
	Erreurs ARCH-ND							
1	5.6	5.4	5.5	5.5	10.3	10.0	10.8	10.8
2	5.0	5.0	5.0	4.9	8.7	8.7	9.9	9.7
3	4.5	4.4	5.0	5.0	8.7	8.5	9.4	9.3
4	4.4	4.3	5.3	5.3	9.2	9.0	11.1	10.9
5	3.3	3.1	5.3	5.3	6.7	6.7	12.3	12.1
6	4.7	4.3	5.7	5.5	8.9	8.5	11.2	10.9
7	4.1	3.7	4.5	4.5	7.3	6.7	9.1	8.8
8	4.1	4.0	5.9	5.8	7.5	7.2	11.1	11.1

Annexe C

CODE INFORMATIQUE

Voici le code pour les erreurs ARCH, n=600.

```
************************
on initialise le nombre d'itération(B), le nombre de délais
individuels(M), le nombre de séries considérées(d) et les valeurs
d'intérêt pour la statistique Portemanteau(taille)
************************
B=1000
M=10
d=2
n=600
taille_vector(length=4)
taille[1]=2
taille[2]=3
taille[3]=6
taille[4]=10
************************
on initialise les points critiques
************************
pcind5_qchisq(0.95,d*d)
pcind10_qchisq(0.90,d*d)
pcind1_qchisq(0.99,d*d)
pcport5_vector(length=(length(taille)))
```

```
pcport10_vector(length=(length(taille)))
pcport1_vector(length=(length(taille)))
pcport5[1]_qchisq(0.95,taille[1]*d*d)
pcport10[1]_qchisq(0.90,taille[1]*d*d)
pcport1[1]_qchisq(0.99,taille[1]*d*d)
pcport5[2]_qchisq(0.95,taille[2]*d*d)
pcport10[2]_qchisq(0.90,taille[2]*d*d)
pcport1[2]_qchisq(0.99,taille[2]*d*d)
pcport5[3]_qchisq(0.95,taille[3]*d*d)
pcport10[3]_qchisq(0.90,taille[3]*d*d)
pcport1[3]_qchisq(0.99,taille[3]*d*d)
pcport5[4]_qchisq(0.95,taille[4]*d*d)
pcport10[4]_qchisq(0.90,taille[4]*d*d)
pcport1[4]_qchisq(0.99,taille[4]*d*d)
***********************
on initialise les vecteurs et matrices utilisés
***********************
resultQMtest_matrix(0,nrow=B,ncol=(length(taille)))
resultQM5<-matrix(0,ncol=1,nrow=(length(taille)))
resultQM10<-matrix(0,ncol=1,nrow=(length(taille)))
resultQM1<-matrix(0,ncol=1,nrow=(length(taille)))
resultQMstar5<-matrix(0,ncol=1,nrow=(length(taille)))
resultQMstar10<-matrix(0,ncol=1,nrow=(length(taille)))
resultQMstar1<-matrix(0,ncol=1,nrow=(length(taille)))
resultQM_vector(length=(length(taille)))
resultQMstar_vector(length=(length(taille)))
resultQLtest_matrix(0,nrow=B,ncol=M)
resultQL5<-matrix(0,ncol=1,nrow=M)
resultQL10<-matrix(0,ncol=1,nrow=M)
resultQL1<-matrix(0,ncol=1,nrow=M)
resultQLstar5<-matrix(0,ncol=1,nrow=M)
```

```
resultQLstar10<-matrix(0,ncol=1,nrow=M)
resultQLstar1<-matrix(0,ncol=1,nrow=M)
resultQl_vector(length=M)
resultQlstar_vector(length=M)
tempQl_vector(length=M)
tempQlstar_vector(length=M)
resultQMtestgen_matrix(0,nrow=B,ncol=(length(taille)))
resultQM5gen<-matrix(0,ncol=1,nrow=(length(taille)))
resultQM10gen<-matrix(0,ncol=1,nrow=(length(taille)))
resultQM1gen<-matrix(0,ncol=1,nrow=(length(taille)))
resultQMstar5gen<-matrix(0,ncol=1,nrow=(length(taille)))
resultQMstar10gen<-matrix(0,ncol=1,nrow=(length(taille)))
resultQMstar1gen<-matrix(0,ncol=1,nrow=(length(taille)))
resultQMgen_vector(length=(length(taille)))
resultQMstargen_vector(length=(length(taille)))
resultQLtestgen_matrix(0,nrow=B,ncol=M)
resultQL5gen<-matrix(0,ncol=1,nrow=M)
resultQL10gen<-matrix(0,ncol=1,nrow=M)
resultQL1gen<-matrix(0,ncol=1,nrow=M)
resultQLstar5gen<-matrix(0,ncol=1,nrow=M)
resultQLstar10gen<-matrix(0,ncol=1,nrow=M)
resultQLstar1gen<-matrix(0,ncol=1,nrow=M)
resultQlgen_vector(length=M)
resultQlstargen_vector(length=M)
tempQlgen_vector(length=M)
tempQlstargen_vector(length=M)
resultQMtestsc_matrix(0,nrow=B,ncol=(length(taille)))
resultQM5sc<-matrix(0,ncol=1,nrow=(length(taille)))
resultQM10sc<-matrix(0,ncol=1,nrow=(length(taille)))
resultQM1sc<-matrix(0,ncol=1,nrow=(length(taille)))
resultQMstar5sc<-matrix(0,ncol=1,nrow=(length(taille)))
```

```
resultQMstar10sc<-matrix(0,ncol=1,nrow=(length(taille)))
resultQMstar1sc<-matrix(0,ncol=1,nrow=(length(taille)))
resultQMsc_vector(length=(length(taille)))
resultQMstarsc_vector(length=(length(taille)))
resultQMtest_matrix(0,nrow=B,ncol=(length(taille)))
I_diag(M)
Yt_matrix(nrow=2,ncol=(n+200))
Ft_vector(length=(n+200))
Fttrans_vector(length=(n+200))
ca_vector(length=M*d*d)
castar_vector(length=M*d*d)
***********************
on initialise les valeurs des paramètres du modèle
***********************
phi111=0.7
phi112=0
phi121=0.3
phi122=0.7
phi211=-0.7
phi212=0
phi221=-0.3
phi222=-0.7
***********************
on crée les fonctions nécessaires pour chaque itération
***********************
phi1<-function(phi111,phi112,phi121,phi122){

phi1<-matrix(nrow=2,ncol=2)
phi1[1,1]=phi111
phi1[1,2]=phi112
phi1[2,1]=phi121
```

```
phi1[2,2]=phi122
phi1
}

vec<-function(x){
y<-t(t(as.vector(x)))
y
}



For(w=1:B,{
***********************
on crée le terme d'erreur
***********************
At1_matrix(0,nrow=(n+1501),ncol=2)

Attemp1_rnorm(1)
Attemp2_rnorm(1)
At1[1,1]_Attemp1
At1[1,2]_Attemp2

for( i in 2:(n+1501)){

At1[i,1]=rnorm(1,mean=0,sd=(sqrt(0.7+0.3*(At1[(i-1),1])^2)))
At1[i,2]=rnorm(1,mean=0,sd=(sqrt(0.5+0.5*(At1[(i-1),2])^2)))

}

At1_At1[1301:(n+1501),]

Yt[1,1]=0.01
```

```
Yt[2,1]=0.01
Fttrans[1]=0
************************
on génère les données
************************
for(i in 2:(n+200)){
Ft=ifelse(Yt[1,i-1]>0,1,0)
Fttrans[i]=ifelse(Yt[1,i-1]>0,1,0)
Yt[1,i]=Ft*(phi111*Yt[1,i-1]+phi112*Yt[2,i-1]
+At1[i,1])+(1-Ft)*(phi211*Yt[1,i-1]+phi212*Yt[2,i-1]+At1[i,1])
Yt[2,i]=Ft*(phi121*Yt[1,i-1]+phi122*Yt[2,i-1]
+At1[i,2])+(1-Ft)*(phi221*Yt[1,i-1]+phi222*Yt[2,i-1]+At1[i,2])
}

Ftstat<-Fttrans[201:(n+200)]
Ytstat<-matrix(nrow=2,ncol=n)
Yt1statreg1<-matrix(nrow=2,ncol=n)
Yt1statreg2<-matrix(nrow=2,ncol=n)


for(i in 1:n){

Ytstat[1,i]=Yt[1,200+i]
Ytstat[2,i]=Yt[2,200+i]
}

for(i in 1:n){
Yt1statreg1[1,i]=Yt[1,199+i]*Ftstat[i]
Yt1statreg1[2,i]=Yt[2,199+i]*Ftstat[i]
}
```

```
for(i in 1:n){
Yt1statreg2[1,i]=Yt[1,199+i]*(1-Ftstat[i])
Yt1statreg2[2,i]=Yt[2,199+i]*(1-Ftstat[i])
}
************************
on estime le modèle
************************
setar<-nls(Ytstat~phi1(phi111,phi112,phi121,phi122)%*%Yt1statreg1
+phi1(phi211,phi212,phi221,phi222)%*%Yt1statreg2,
start=list(phi111=0.7,phi112=0.01,phi121=0.3,phi122=0.7,
phi211=-0.69,phi212=-0.01,phi221=-0.29,phi222=-0.69))

phi111hat=setar[1]$parameters[1]
phi112hat=setar[1]$parameters[2]
phi121hat=setar[1]$parameters[3]
phi122hat=setar[1]$parameters[4]

phi211hat=setar[1]$parameters[5]
phi212hat=setar[1]$parameters[6]
phi221hat=setar[1]$parameters[7]
phi222hat=setar[1]$parameters[8]

A<-function(t){

A<-matrix(nrow=8,ncol=2)

A[1,1]=Yt1statreg1[1,t]
A[2,1]=Yt1statreg1[2,t]
A[3,1]=0
A[4,1]=0
A[5,1]=Yt1statreg2[1,t]
```

```
A[6,1]=Yt1statreg2[2,t]
A[7,1]=0
A[8,1]=0

A[1,2]=0
A[2,2]=0
A[3,2]=Yt1statreg1[1,t]
A[4,2]=Yt1statreg1[2,t]
A[5,2]=0
A[6,2]=0
A[7,2]=Yt1statreg2[1,t]
A[8,2]=Yt1statreg2[2,t]

A

}
***********************
on garde les résidues
***********************
ressetar<-Ytstat-(phi1(phi111hat,phi112hat,phi121hat,phi122hat)
%*%Yt1statreg1+phi1(phi211hat,phi212hat,phi221hat,phi222hat)
%*%Yt1statreg2)

reshat_t(ressetar)
***********************
on calcule les matrices pour la statistique de test
***********************
sigmahat<-(1/n)*(t(reshat)%*%reshat)

sigmahatinverse_solve(sigmahat)
***********************
```

```
la matrice U
***********************
Uhat<-matrix(0,nrow=8,ncol=8)

for(i in 1:n){

Uhat=Uhat+((1/n)*(A(i)%*%sigmahatinverse%*%t(A(i))))

}
***********************
la matrice R
***********************
Rhat<-matrix(0,nrow=8,ncol=8)

for( i in 1:n){
Rhat_Rhat+((1/n)*((A(i)%*%sigmahatinverse%*%reshat[i,])
%*%(t(reshat[i,])%*%sigmahatinverse%*%t(A(i)))))
}

Var<-solve(Uhat)

***********************
la matrice J
***********************
Jhat<-matrix(0,nrow=M*d*d,ncol=8)

 for( i in 1:M){

 J1hat<-matrix(0,nrow=8,ncol=d*d)

 for(k in (i+1):n){
```

```
J1hat<-J1hat+kronecker(t(reshat[k-i,]),A(k))
}

J1hat<-(1/n)*J1hat
Jhat[(d*d*(i-1)+1):(d*d*i),1:8]=(t(J1hat))

}
***********************
la matrice Jtild
***********************
Jhattild<-matrix(0,nrow=M*d*d,ncol=8)

for( i in 1:M){

J1hattild<-matrix(0,nrow=8,ncol=d*d)

for(k in (i+1):n){

temp2_(A(k)%*%(sigmahatinverse%*%reshat[k,]%*%t(reshat[k,])))

J1hattild<-J1hattild+kronecker(t(reshat[k-i,]),temp2)
}

J1hattild<-(1/n)*J1hattild

Jhattild[(d*d*(i-1)+1):(d*d*i),1:8]=t(J1hattild)

}
***********************
on estime les autocovariances résiduelles
```

```
***********************
for(i in 1:M){

ca1<-matrix(0,nrow=d,ncol=d)

for(k in (i+1):n){

ca1<-ca1+(t(t(reshat[k,])))%*%(t(reshat[(k-i),]))

}

ca[(d*d*(i-1)+1):(d*d*i)]<-vec(ca1)

}

ca<-(1/n)*vec(ca)

for(i in 1:M){

cal=ca[(d*d*(i-1)+1):(d*d*(i-1)+4)]
***********************
on calcule les statistiques pour les  délais individuels
***********************
Omegal<-kronecker(sigmahat,sigmahat)
-Jhat[(d*d*(i-1)+1):(d*d*i),1:8]
%*%Var%*%(t(Jhat[(d*d*(i-1)+1):(d*d*i),1:8]))

resultQl[i]<-n*(t(cal))%*%(solve(Omegal))%*%(t(t(cal)))
resultQlstar[i]<-(n/(n-i))*resultQl[i]
```

```
resultQLtest[w,i]=resultQl[i]

resultQL5[i]<-resultQL5[i]
+ifelse(resultQl[i]>pcind5,1,0)
resultQL10[i]<-resultQL10[i]
+ifelse(resultQl[i]>pcind10,1,0)
resultQL1[i]<-resultQL1[i]
+ifelse(resultQl[i]>pcind1,1,0)

resultQLstar5[i]<-resultQLstar5[i]
+ifelse(resultQlstar[i]>pcind5,1,0)
resultQLstar10[i]<-resultQLstar10[i]
+ifelse(resultQlstar[i]>pcind10,1,0)
resultQLstar1[i]<-resultQLstar1[i]
+ifelse(resultQlstar[i]>pcind1,1,0)

Varcal_matrix(0,nrow=d*d,ncol=d*d)

for( j in (M+1):n)

{

Varcal_Varcal+kronecker(reshat[(j-i),]
%*%t(reshat[(j-i),]),reshat[j,]%*%t(reshat[j,]))

}

Varcal=(1/n)*Varcal

Omegalgen_Varcal-Jhat[(d*d*(i-1)+1):(d*d*i),1:8]
%*%Var%*%(t(Jhattild[(d*d*(i-1)+1):(d*d*i),1:8]))
```

```
-Jhattild[(d*d*(i-1)+1):(d*d*i),1:8]%*%
Var%*%(t(Jhat[(d*d*(i-1)+1):(d*d*i),1:8]))
+Jhat[(d*d*(i-1)+1):(d*d*i),1:8]
%*%(Var%*%Rhat%*%Var)%*%(t(Jhat[(d*d*(i-1)+1):(d*d*i),1:8]))

resultQlgen[i]<-n*(t(cal))%*%(solve(Omegalgen))%*%(t(t(cal)))
resultQlstargen[i]<-(n/(n-i))*resultQlgen[i]

resultQLtestgen[w,i]=resultQlgen[i]

resultQL5gen[i]<-resultQL5gen[i]
+ifelse(resultQlgen[i]>pcind5,1,0)
resultQL10gen[i]<-resultQL10gen[i]
+ifelse(resultQlgen[i]>pcind10,1,0)
resultQL1gen[i]<-resultQL1gen[i]
+ifelse(resultQlgen[i]>pcind1,1,0)

resultQLstar5gen[i]<-resultQLstar5gen[i]
+ifelse(resultQlstargen[i]>pcind5,1,0)
resultQLstar10gen[i]<-resultQLstar10gen[i]
+ifelse(resultQlstargen[i]>pcind10,1,0)
resultQLstar1gen[i]<-resultQLstar1gen[i]
+ifelse(resultQlstargen[i]>pcind1,1,0)

}

for(i in 1:M){

castar1<-matrix(0,nrow=d,ncol=d)

for(k in (i+1):n){
```

```
castar1<-castar1+(t(t(reshat[k,])))%*%(t(reshat[(k-i),]))

}

castar[(d*d*(i-1)+1):(d*d*i)]<-sqrt((n/(n-i)))*vec(castar1)

}

castar<-(1/n)*vec(castar)

astack_function(t,reshat,M,d){

result_matrix(0,nrow=d*M,ncol=1)

for(j in 1:M){

result[(2*j-1):(2*j)]=vec(reshat[(M+t-j),])
}
result
}

Varca_matrix(0,nrow=M*d*d,ncol=M*d*d)

for( j in 1:(n-M))

{

Varca_Varca+kronecker(astack(j,reshat,M,d)
%*%t(astack(j,reshat,M,d)),reshat[(j+M),]%*%t(reshat[j+M,]))
}
```

```
Varca=(1/n)*Varca

for(i in 1:(length(taille))){

catemp_ca[1:(d*d*taille[i])]

castartemp_castar[1:(d*d*taille[i])]
************************
on calcule les statistiques portemanteaux
************************
Omega<-kronecker(kronecker(diag(taille[i]),sigmahat),sigmahat)
-Jhat[1:(d*d*taille[i]),1:8]
%*%Var%*%(t(Jhat[1:(d*d*taille[i]),1:8]))

Omega<-solve(Omega)

Omegagen<-Varca[1:(d*d*taille[i]),1:(d*d*taille[i])]
-Jhattild[1:(d*d*taille[i]),1:8]%*%
Var%*%(t(Jhat[1:(d*d*taille[i]),1:8]))
-Jhat[1:(d*d*taille[i]),1:8]%*%Var%*%
(t(Jhattild[1:(d*d*taille[i]),1:8]))+Jhat[1:(d*d*taille[i]),1:8]
%*%(Var%*%Rhat%*%Var)
%*%(t(Jhat[1:(d*d*taille[i]),1:8]))

Omegagen<-solve(Omegagen)

Omegasc<-kronecker(kronecker(diag(taille[i]),sigmahat),sigmahat)
Omegasc<-solve(Omegasc)

resultQM[i]=n*t(catemp)%*%Omega%*%catemp
```

```
resultQMstar[i]=n*t(castartemp)%*%Omega%*%castartemp
resultQMgen[i]=n*t(catemp)%*%Omegagen%*%catemp
resultQMstargen[i]=n*t(castartemp)%*%Omegagen%*%castartemp
resultQMsc[i]=n*t(catemp)%*%Omegasc%*%catemp
resultQMstarsc[i]=n*t(castartemp)%*%Omegasc%*%castartemp

resultQMtest[w,i]=resultQM[i]
resultQMtestgen[w,i]=resultQMgen[i]

resultQM5[i]<-resultQM5[i]
+ifelse(resultQM[i]>pcport5[i],1,0)
resultQM10[i]<-resultQM10[i]
+ifelse(resultQM[i]>pcport10[i],1,0)
resultQM1[i]<-resultQM1[i]
+ifelse(resultQM[i]>pcport1[i],1,0)

resultQMstar5[i]<-resultQMstar5[i]
+ifelse(resultQMstar[i]>pcport5[i],1,0)
resultQMstar10[i]<-resultQMstar10[i]
+ifelse(resultQMstar[i]>pcport10[i],1,0)
resultQMstar1[i]<-resultQMstar1[i]
+ifelse(resultQMstar[i]>pcport1[i],1,0)


resultQM5gen[i]<-resultQM5gen[i]
+ifelse(resultQMgen[i]>pcport5[i],1,0)
resultQM10gen[i]<-resultQM10gen[i]
+ifelse(resultQMgen[i]>pcport10[i],1,0)
resultQM1gen[i]<-resultQM1gen[i]
+ifelse(resultQMgen[i]>pcport1[i],1,0)
```

```
resultQMstar5gen[i]<-resultQMstar5gen[i]
+ifelse(resultQMstargen[i]>pcport5[i],1,0)
resultQMstar10gen[i]<-resultQMstar10gen[i]
+ifelse(resultQMstargen[i]>pcport10[i],1,0)
resultQMstar1gen[i]<-resultQMstar1gen[i]
+ifelse(resultQMstargen[i]>pcport1[i],1,0)

resultQM5sc[i]<-resultQM5sc[i]
+ifelse(resultQMsc[i]>pcport5[i],1,0)
resultQM10sc[i]<-resultQM10sc[i]
+ifelse(resultQMsc[i]>pcport10[i],1,0)
resultQM1sc[i]<-resultQM1sc[i]
+ifelse(resultQMsc[i]>pcport1[i],1,0)

resultQMstar5sc[i]<-resultQMstar5sc[i]
+ifelse(resultQMstarsc[i]>pcport5[i],1,0)
resultQMstar10sc[i]<-resultQMstar10sc[i]
+ifelse(resultQMstarsc[i]>pcport10[i],1,0)
resultQMstar1sc[i]<-resultQMstar1sc[i]
+ifelse(resultQMstarsc[i]>pcport1[i],1,0)
}
},grain.size=5)
************************
on imprime les résultats
************************
Resultdel_cbind(resultQL5,resultQL10,resultQL1
,resultQLstar5,resultQLstar10,resultQLstar1)

Resultport_cbind(resultQM5,resultQM10,resultQM1,resultQMstar5
,resultQMstar10,resultQMstar1,resultQM5sc,resultQM10sc,
resultQM1sc,resultQMstar5sc,resultQMstar10sc,resultQMstar1sc)
```

```
Resultdelgen_cbind(resultQL5gen,resultQL10gen,resultQL1gen
,resultQLstar5gen,resultQLstar10gen,resultQLstar1gen)

Resultportgen_cbind(resultQM5gen,resultQM10gen,resultQM1gen
,resultQMstar5gen,resultQMstar10gen,resultQMstar1gen)

Resultdel
Resultport
Resultdelgen
Resultportgen
}
```

BIBLIOGRAPHIE

Bollerslev, T. (1986), 'Generalized autoregressive conditional heteroskedasticity', *Journal of Econometrics* **3**, 307-327.

Box, G. E. P., Jenkins, G. M. et Reinsel, G. C. (1994), *Time Series Analysis : Forecasting and Control*, 3e édition, Prentice Hall : New-Jersey.

Brockwell, P. J. et Davis, R. A. (2002), *Introduction to Time Series and Forecasting*, 2e édition, Springer : New York.

Chan, K. S. et Tong, H. (1986), 'On estimating thresholds in autoregressive models', *Journal of Time Series Analysis* **7**, 179-190.

Duchesne, P. (2004), 'On matricial measures of dependence in vector ARCH models with applications to diagnostic checking', *Statistics and Probability Letters* **68**, 149-160.

Duchesne, P. (2005), 'On the asymptotic distribution of residual autocovariances in VARX models with applications', *Test* **14**, 449-473.

Engle, R. (1982), 'Autoregressive conditionnal heteroscedasticity with estimates of the United Kingdom inflation', *Econometrica* **50**, 987-1007.

Engle, R. et Kroner, K. F. (1995), 'Multivariate simultaneous generalized ARCH', *Econometric Theory* **11**, 122-150.

Gallant, R. A. (1987), *Nonlinear Statistical Models*, Wiley Series in Probability and Mathematical Statistics, Wiley : New York.

Granger, C. W. J. et Andersen, A. P. (1978), *An Introduction to Bilinear Time Series Models*, Gottingen : Vandenhack and Ruprecht.

Granger, C. W. J. et Teräsvirta, T. (1993), *Modelling Nonlinear Economic Relationships*, Advanced Texts in Econometrics, Oxford University Press : Oxford.

Hannan, E. J. (1970), *Multiple Time Series*, Wiley : New York.

Harville, D. A. (1997), *Matrix Algebra from a Statistician's Perspective*, Springer-Verlag : New York.

Hosking, J. (1980), 'The multivariate portmanteau statistic', *Journal of the American Statistical Association* **75**, 602-608.

Ibragimov, I. A. (1963), 'A central limit theorem for a class of dependant random variables', *Theory of Probability and Its Applications* **8**, 83-89.

Klimko, L. A. et Nelson, P. I. (1978), 'On the conditional least squares estimation for stochastic processes', *Annals of Statistics* **6**, 629-642.

Lasota, A. et Mackey, M. C. (1989), 'Stochastic perturbation of dynamical systems : the weak convergences of measures', *Journal of Mathematics Analysis and Applications* **138**, 232-248.

Li, W. K. (2004), *Diagnostic Checks in Time Series*, Chapman & Hall/CRC : New York.

Li, W. K. (1992), 'On the asymptotic standard errors of residual autocorrelations in nonlinear time series modelling', *Biometrika* **79**, 435-437.

Li, W. K. et McLeod, A. I. (1981), 'Distribution of the residual autocorrelations in multivariate ARMA time series models', *Journal of the Royal Statistical Society, Series B* **43**, 231-239.

Ljung, G. M. et Box, G. E. P. (1978), 'On a measure of lack of fit in time series models', *Biometrika* **65**, 297-304.

Lütkepohl, H. (2005), *New Introduction to Multiple Time Series Analysis*, Springer-Verlag : Berlin.

Stout, W. F. (1974), *Almost Sure Convergence*, Academic Press : New-York.

Taniguchi, M. et Kakizawa, Y. (2000), *Asymptotic Theory of Statistical Inference for Time Series*, Springer-Verlag : New York .

Teräsvirta, T. (1994), 'Specification, estimation and evaluation of smooth transition autoregressive models', *Journal of the American Statistical Association* **89**, 208-218.

Tjøstheim, D. (1986), 'Estimation in nonlinear time series models', *Stochastic Process and their Applications* **21**, 251-273.

Tong, H. (1978), 'On a threshold model', in *Pattern Recognition and Signal Processing*, C. H. Chen (Ed.), Sijtoff and Noordhoff : Amsterdam.

Tong, H. (1983), *Threshold Models in Non-Linear Time Series Analysis*, Springer-Verlag : New York.

Tong, H. (1990), *Non-Linear Time Series : A Dynamical System Approach*, Oxford University Press : Oxford.

Tsay, R. S. (1998), 'Testing and modeling multivariate threshold models', *Journal of the American Statistical Association* **93**, 1188-1202.

van Dijk, D., Teräsvirta, T. et Franses, H. P. (2002), 'Smooth transition autoregressive models – a survey of recent developments', *Econometric Reviews* **21**, 1-47.