

Université de Montréal

**Sur les tests lisses d'ajustement dans le contexte
des séries chronologiques**

par

Joseph François Tagne Tatsinkou

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

Thèse présentée à la Faculté des études supérieures
en vue de l'obtention du grade de
Philosophiæ Doctor (Ph.D.)
en statistique

02 Décembre 2015

© Joseph François Tagne Tatsinkou, 2015

Université de Montréal

Faculté des études supérieures

Cette thèse intitulée

Sur les tests lisses d'ajustement dans le contexte des séries chronologiques

présentée par

Joseph François Tagne Tatsinkou

a été évaluée par un jury composé des personnes suivantes :

Jean-François Angers

(président-rapporteur)

Pierre Duchesne

(directeur de recherche)

Pierre Lafaye de Micheaux

(codirecteur)

Mylène Bédard

(membre du jury)

Alain Latour

(examineur externe)

Marine Carrasco

(représentant du doyen de la FAS)

Thèse acceptée le

21 Avril 2016

DÉDICACES

À la mémoire de mon père **Donatien TATSINKOU**
décédé le 31 Janvier 2004.

À **Vanessa**

SOMMAIRE

La plupart des modèles en statistique classique repose sur une hypothèse sur la distribution des données ou sur une distribution sous-jacente aux données. La validité de cette hypothèse permet de faire de l'inférence, de construire des intervalles de confiance ou encore de tester la fiabilité du modèle. La problématique des tests d'ajustement vise à s'assurer de la conformité ou de la cohérence de l'hypothèse avec les données disponibles. Dans la présente thèse, nous proposons des tests d'ajustement à la loi normale dans le cadre des séries chronologiques univariées et vectorielles. Nous nous sommes limités à une classe de séries chronologiques linéaires, à savoir les modèles autorégressifs à moyenne mobile (ARMA ou VARMA dans le cas vectoriel).

Dans un premier temps, au cas univarié, nous proposons une généralisation du travail de Ducharme et Lafaye de Micheaux (2004) dans le cas où la moyenne est inconnue et estimée. Nous avons estimé les paramètres par une méthode rarement utilisée dans la littérature et pourtant asymptotiquement efficace. En effet, nous avons rigoureusement montré que l'estimateur proposé par Brockwell et Davis (1991, section 10.8) converge presque sûrement vers la vraie valeur inconnue du paramètre. De plus, nous fournissons une preuve rigoureuse de l'inversibilité de la matrice des variances et des covariances de la statistique de test à partir de certaines propriétés d'algèbre linéaire. Le résultat s'applique aussi au cas où la moyenne est supposée connue et égale à zéro. Enfin, nous proposons une méthode de sélection de la dimension de la famille d'alternatives de type AIC, et nous étudions les propriétés asymptotiques de cette méthode. L'outil proposé ici est basé sur une famille spécifique de polynômes orthogonaux, à savoir les polynômes de Legendre.

Dans un second temps, dans le cas vectoriel, nous proposons un test d'ajustement pour les modèles autorégressifs à moyenne mobile avec une paramétrisation structurée. La paramétrisation structurée permet de réduire le nombre élevé de

paramètres dans ces modèles ou encore de tenir compte de certaines contraintes particulières. Ce projet inclut le cas standard d'absence de paramétrisation. Le test que nous proposons s'applique à une famille quelconque de fonctions orthogonales. Nous illustrons cela dans le cas particulier des polynômes de Legendre et d'Hermite. Dans le cas particulier des polynômes d'Hermite, nous montrons que le test obtenu est invariant aux transformations affines et qu'il est en fait une généralisation de nombreux tests existants dans la littérature. Ce projet peut être vu comme une généralisation du premier dans trois directions, notamment le passage de l'univarié au multivarié ; le choix d'une famille quelconque de fonctions orthogonales ; et enfin la possibilité de spécifier des relations ou des contraintes dans la formulation VARMA.

Nous avons procédé dans chacun des projets à une étude de simulation afin d'évaluer le niveau et la puissance des tests proposés ainsi que de les comparer aux tests existants. De plus des applications aux données réelles sont fournies. Nous avons appliqué les tests à la prévision de la température moyenne annuelle du globe terrestre (univarié), ainsi qu'aux données relatives au marché du travail canadien (bivarié).

Ces travaux ont été exposés à plusieurs congrès (voir par exemple Tagne, Duchesne et Lafaye de Micheaux (2013a, 2013b, 2014) pour plus de détails). Un article basé sur le premier projet est également soumis dans une revue avec comité de lecture (Voir Duchesne, Lafaye de Micheaux et Tagne (2016)).

Mots clés : Test d'ajustement, test lisse de Neyman, polynômes orthogonaux, séries chronologiques, hypothèse de normalité, modèle ARMA, modèle VARMA.

SUMMARY

Several phenomena from natural and social sciences rely on distribution's assumption among which the normal distribution is the most popular. The validity of that assumption is useful to setting up forecast intervals or for checking model adequacy of the underlying model. The goodness-of-fit procedures are tools to assess the adequacy of the data's underlying assumptions. Autoregressive and moving average time series models are often used to find the mathematical behavior of these phenomena from natural and social sciences, and especially in the finance area. These models are based on some assumptions including normality distribution for the innovations. Normality assumption may be helpful for some testing procedures. Furthermore, stronger conclusions can be drawn from the adjusted model if the white noise can be assumed Gaussian. In this work, goodness-of-fit tests for checking normality for the innovations from autoregressive moving average time series models are proposed for both univariate and multivariate cases (ARMA and VARMA models).

In our first project, a smooth test of normality for ARMA time series models with unknown mean based on a least square type estimator is proposed. We derive the asymptotic null distribution of the test statistic. The result here is an extension of the paper of Ducharme et Lafaye de Micheaux (2004), where they supposed the mean known and equal to zero. We use the least square type estimator proposed by Brockwell et Davis (1991, section 10.8) and we provide a rigorous proof that it is almost surely convergent. We show that the covariance matrix of the test is nonsingular regardless if the mean is known. We have also studied a data driven approach for the choice of the dimension of the family and we gave a finite sample approximation of the null distribution. Finally, the finite and asymptotic sample properties of the proposed test statistic are studied via a small simulation study.

X

In the second project, goodness-of-fit tests for checking multivariate normality for the innovations from vector autoregressive moving average time series models are proposed. Since these time series models may rely on a large number of parameters, structured parameterization of the functional form is allowed. The methodology also relies on the smooth test paradigm and on families of orthonormal functions with respect to the multivariate normal density. It is shown that the smooth tests converge to convenient chi-square distributions asymptotically. An important special case makes use of Hermite polynomials, and in that situation we demonstrate that the tests are invariant under linear transformations. We observed that the test is not invariant under linear transformations with Legendre polynomials. A consistent data driven method is discussed to choose the family order from the data. In a simulation study, exact levels are studied and the empirical powers of the smooth tests are compared to those of other methods. Finally, an application to real data is provided, specifically on Canadian labour market data and annual global temperature.

These works were exposed at several meeting (see for example Tagne, Duchesne and Lafaye de Micheaux (2013a, 2013b, 2014) for more details). A paper based on the first project is submitted in a refereed journal (see Duchesne, Lafaye de Micheaux et Tagne (2016)).

Keywords : Goodness-of-fit test, Neyman's smooth test, orthogonal polynomials, time series, normality test, ARMA model, VARMA model.

TABLE DES MATIÈRES

Dédicaces.....	v
Sommaire.....	vii
Summary.....	ix
Liste des figures	xv
Liste des tableaux	xvii
Quelques notations.....	xix
Quelques notations.....	xix
Remerciements	xxi
Introduction générale.....	1
Problématique.....	1
Structure de la thèse.....	3
Chapitre 1. Introduction aux modèles de séries chronologiques..	5
1.1. Processus stationnaires, linéaires et inversibles	6
1.1.1. Cas univarié	6
1.1.2. Cas vectoriel.....	9
1.2. La classe des modèles ARMA	11
1.2.1. Formulation et propriétés	11
1.2.2. Estimation des paramètres	13
1.3. La classe des modèles ARMA vectoriels	18
1.3.1. Spécification	18
1.3.2. Estimation d'un modèle VARMA	19
1.3.3. Paramétrisation structurée d'un VARMA	21
Chapitre 2. Introduction aux tests lisses d'ajustement	25

2.1.	Quelques notions d'algèbre linéaire et d'analyse fonctionnelle	26
2.2.	Le principe du test lisse de Neyman	30
2.2.1.	Le test du score de Rao	30
2.2.2.	Construction du test lisse de Neyman	31
2.2.3.	Obtention de la statistique de test	32
2.2.4.	Le choix de la dimension de la famille d'alternatives	34
Chapitre 3. Test lisse d'ajustement pour les modèles ARMA avec moyenne inconnue		37
3.1.	Introduction	37
3.2.	Étude de la convergence forte	42
3.3.	Étude de la normalité asymptotique de l'estimateur ν^\dagger	50
3.4.	Étude de la normalité asymptotique du test lisse de normalité	51
3.5.	Choix de K	54
3.6.	Statistique alternative reposant sur les résidus centrés	56
3.7.	Quelques évidences empiriques	59
3.8.	Conclusion	64
Chapitre 4. Tests lisses d'ajustement pour les séries chronologiques multivariées (VARMA)		65
4.1.	Introduction et préliminaires	66
4.1.1.	Introduction	66
4.1.2.	Préliminaires	69
4.2.	Tests lisses pour un modèle de séries chronologiques VARMA structuré	72
4.2.1.	Le cas des polynômes d'Hermitte normalisés	77
4.2.2.	Le cas des polynômes de Legendre	79
4.3.	Sélection automatique de l'ordre R	81
4.4.	Simulation numérique	83
4.5.	Une application aux données macroéconomiques canadiennes	94
4.6.	Conclusion	96

Conclusion générale	97
Annexe A. Compléments du chapitre 3	A-i
A.1. Détails des principales preuves.....	A-i
A.1.1. Preuve de (3.4.2).....	A-i
A.1.2. Preuve de la proposition 3.3.1.....	A-iii
A.1.3. Preuve de (3.4.3).....	A-vi
A.1.4. Preuve de (3.4.5).....	A-viii
A.1.5. Preuve de (3.5.2).....	A-ix
A.1.6. Preuve du théorème 3.6.1.....	A-xi
Annexe B. Compléments du chapitre 4	B-i
B.1. Preuve des principaux résultats.....	B-i
B.1.1. Preuve du théorème 4.2.1.....	B-ii
B.1.2. Preuve du lemme 4.2.1.....	B-viii
Illustration du lemme B.1.2 pour $m = 2$, $k = 1$ et $k = 2$	B-ix
Preuve du lemme 4.2.1.....	B-ix
B.1.3. Preuve du théorème 4.2.2.....	B-ix
B.2. Tableaux et figures supplémentaires.....	B-xi
Annexe C. Code Informatique	C-i
Références bibliographiques	C-i

LISTE DES FIGURES

1.1	Déviations de la température moyenne annuelle (en Celsius) du globe terrestre entre 1880 et 2009.	6
4.1	Séries chronologiques du marché du travail canadien.	94
B.1	Diagnostic du rejet du test lisse avec les polynômes d’Hermite.	B-xv

LISTE DES TABLEAUX

3.1	Quelques quantiles obtenus de l'approximation (3.5.2) avec $d = 2$ pour $\alpha = 1\%$, 5% , 10% et 20%	56
3.2	Niveau empirique (en pourcentage) des différents tests avec 10 000 répliques de Monte Carlo au niveau nominal de 5% pour $n = 200$ et $n = 400$	61
3.3	Puissance empirique (en pourcentage) des différents tests avec 10 000 répliques de Monte Carlo au niveau nominal de 5% pour $n = 50$	63
4.1	Indices multiples ordonnés pour $m = 4$ et $k = 3$	73
4.2	Paramètres des modèles de séries chronologiques DGP_1 , DGP_2 , DGP_3 et DGP_4	84
4.3	Matrices des covariances du bruit dans les modèles DGP_1 , DGP_2 , DGP_3 et DGP_4	84
4.4	Niveaux empiriques (en pourcentage) des différents tests basés sur 1000 répliques au niveau nominal de 5%	92
4.5	Puissances empiriques (en pourcentage) des différents tests basées sur 1000 répliques au niveau nominal de 5% pour $n = 50$ avec les valeurs critiques asymptotiques.	93
4.6	Valeur- p des tests de normalité appliqués aux données macroéconomiques canadiennes.	95
B.1	Puissances empiriques (en pourcentage) des différents tests basées sur 1000 répliques pour un niveau nominal de 5% et pour $n = 50$ avec les valeurs critiques asymptotiques.	B-xii
B.2	Puissances empiriques (en pourcentage) des différents tests basées sur 1000 répliques pour un niveau nominal de 5% et pour $n = 50$ avec les valeurs critiques simulées.	B-xiii

B.3	Puissances empiriques (en pourcentage) des différents tests basées sur 1000 répliques pour un niveau nominal de 5% et pour $n = 50$ avec les valeurs critiques simulées.	B-xiv
B.4	Quelques résultats d'estimation de la modélisation du marché du travail canadien.	B-xv

Quelques notations

Symbole	Signification
$\mathcal{N}(\mu, \sigma)$: Loi normale univariée de moyenne μ et de variance σ^2 .
$\mathcal{N}_d(\boldsymbol{\alpha}, \boldsymbol{\Sigma})$: Loi normale multivariée à d -dimensions de moyenne $\boldsymbol{\alpha}$ et de variance $\boldsymbol{\Sigma}$.
\xrightarrow{d}	: Convergence en loi.
\xrightarrow{p}	: Convergence en probabilité.
$\xrightarrow{p.s.}$: Convergence presque sûre.
$X_n = o_p(a_n)$: La suite X_n/a_n converge en probabilité vers 0.
$X_n = O_p(a_n)$: La suite X_n/a_n est bornée en probabilité.
δ_{ij}	: Fonction de Kronecker.
AR(p)	: Processus autorégressif d'ordre p .
MA(q)	: Processus moyenne mobile d'ordre q .
ARMA(p, q)	: Processus autorégressif et à moyenne mobile d'ordre p et q .
\triangle	: Fin de définition.
\square	: Fin de démonstration.
\blacksquare	: Fin de théorème, proposition, lemme, résultat et autres.
\blacksquare	: Fin d'exemple et de remarque.

Certaines notations liées aux développements mathématiques sont spécifiques à chaque chapitre.

REMERCIEMENTS

L'aboutissement du présent travail n'a été possible que grâce à l'apport multiforme de tous ceux et toutes celles qui ont contribué à son élaboration. Qu'il me soit donné ici l'occasion de leur adresser mes sincères remerciements.

À mes directeurs de recherche. Je remercie honnêtement Pierre Duchesne, mon directeur de recherche. Son amour et sa passion pour ce qu'il fait m'ont largement aidé pendant mon cheminement de l'admission à ce moment. Je remercie particulièrement mon co-directeur de recherche, Pierre Lafaye de Micheaux pour l'implication personnelle et la proximité malgré certaines circonstances. Je ne cacherai pas le privilège que j'ai eu d'être encadré par vous.

Je remercie très humblement le professeur Alain Latour pour avoir accepté de consacrer une partie de son précieux temps à l'examen de cette thèse. Dans le même élan, j'exprime ma reconnaissance aux professeurs Jean-François Angers et Martin Bilodeau pour les idées constructives et les remarques pertinentes proposées lors de la présentation de mon projet de recherche, et finalement pour leur participation dans mon jury de thèse. J'exprime ma sincère reconnaissance à la professeure Mylène Bédard. J'ai pu recevoir d'elle des commentaires rigoureux et pertinents. Je remercie la professeure Marine Carrasco d'avoir accepté d'être témoin du bon déroulement académique de ma défense de thèse.

Ma sincère reconnaissance aux personnels administratifs ainsi qu'au corps professoral du Département de Mathématiques et Statistique (DMS) de l'Université de Montréal pour l'accueil et la sympathie. Je citerai à titre d'illustration tous ceux qui m'ont enseigné à savoir François Perron, Alejandro Murua, Christian Léger, Miguel Chagnon et Sabin Lessard. Je pense particulièrement à Anne-Marie Dupuis pour l'excellente gestion de mon dossier étudiant, ainsi qu'à Émilie Langlois-Dubois. Je remercie le Centre de Recherches Mathématiques (CRM) pour le soutien financier sans lequel cette thèse n'aurait pu être menée à terme. Je dis merci à tous mes camarades du DMS (Aurélien, Joel, Ajavon, Philippe, Thierry, Valéry, Fabiola, Romain, Janie, Bianca, Thanina, Hervé, Romuald, etc) pour ces nombreuses années de travail commun.

Je remercie très sincèrement Paul Gilbert, retraité de la Banque du Canada pour l'échange intime à propos de son package *R*, *dse* qui m'a été d'une grande utilité. J'exprime ma gratitude à tous ceux qui ont accepté volontairement de relire cette thèse. Mohammed Hadj Djeloul, Francis Bomba, Anthony Amichia, Vanessa Dongmo, Aurélien Guetsop.

Il m'est impossible de citer tous ceux qui ont apportés leur aide à ce travail. A tous les amis et camarades à travers le monde, je dis ma très sincère reconnaissance. Je fais ainsi un clin d'oeil à tous les diplômés du réseau des écoles de statistiques africaines et particulièrement à ceux de l'Institut Sous-régional de Statistique et d'Économie Appliquée (ISSEA) de Yaoundé. Je pense aussi à mes amis Cyril et Monique, Hugues et Sonia, Herman et Audrey ; les familles Tchouaket, Njoumené, Dongmo Jiongo, Momeya, Guetsop, Sango.

Je terminerai en faisant un clin d'oeil à ma très chère famille se situant à **9100 km** de Montréal, et plus exactement dans ma terre natale qu'est le Cameroun. Le chagrin que vous portez de me voir loin de vous est incommensurable. Je pense ici à ma très chère maman Nyadjou Pauline, Maman Marie Mbezele, Clovis Djike, Pierre Njiké, Line Leundjeu, Honorine Njiké, Francis Bomba, etc.

INTRODUCTION GÉNÉRALE

PROBLÉMATIQUE

Plusieurs modèles de séries chronologiques reposent sur certaines hypothèses concernant la distribution des erreurs. La validité de ces hypothèses est souvent critiquée lorsque l'on veut faire de l'inférence ou de la prévision dans ces modèles (voir par exemple Andrews (1993)). Comme dans les modèles linéaires, l'hypothèse de normalité est souvent présente. Ainsi, il devient pertinent de considérer le problème du test d'adéquation à une loi normale du terme d'erreur.

Dans un contexte de séries chronologiques, l'hypothèse de normalité facilite l'étude des propriétés des intervalles de prévisions, voir Lütkepohl (2005). Le rejet de cette hypothèse peut indiquer un mauvais ajustement dans certains modèles autorégressifs à moyenne mobile (ARMA), voir par exemple Pierce (1985) ou aussi Brockett, Hinich et Patterson (1988). De plus, ce rejet peut permettre de choisir entre une modélisation gaussienne et une modélisation non gaussienne (voir Li et McLeod (1988)). Comme discuté dans Ducharme et Lafaye de Micheaux (2004), les auteurs qui se sont penchés sur ce problème se divisent principalement en deux approches. La première consiste à tester les observations originales (voir par exemple Lomnicki (1961), Hinick (1982) et Epps (1987)) alors que la seconde consiste à tester les résidus du modèle estimé (voir par exemple Lütkepohl (2005)). Sous certaines conditions, Lütkepohl (2005) mentionne que dans le cadre des modèles autorégressifs univariés et multivariés, les tests basés sur la série chronologique originale affichent des puissances moindres que ceux reposant sur les résidus estimés comme différence entre la série et sa valeur prédite. Ceci suggère qu'il peut être pertinent d'adopter une stratégie où dans un premier temps un modèle est postulé, et dans une seconde étape il est vérifié si les erreurs (composantes aléatoires non prévisibles du modèle) sont compatibles avec une hypothèse de normalité.

Les tests basés sur les résidus impliquent une attention particulière relativement à la distribution asymptotique de la statistique de test. En effet, il est

2

tendant d'adapter les méthodologies existantes dans un contexte de données dépendantes. Cependant, on ne dispose plus d'observations indépendantes et identiquement distribuées (i.i.d.) comme dans la plupart des tests standards et les valeurs critiques de ces tests ne sont plus toujours appropriées. Quelques auteurs ont regardé l'effet de l'utilisation des résidus sur certains tests d'ajustement dans le contexte des séries chronologiques. Pierce (1985) montre que tout test de normalité calculé à partir des résidus des modèles autorégressifs possède la même distribution asymptotique que dans les cas standards d'observations i.i.d. avec moyenne et variance estimées. On trouve donc que l'estimation des paramètres dans les modèles autorégressifs n'a pas d'effet asymptotiquement. Ce résultat est généralisé dans le cadre des modèles vectoriels autorégressifs (VAR) par Lütkepohl (2005, p. 178), qui montre que la distribution asymptotique de la statistique de test basée sur les coefficients d'asymétrie et d'aplatissement reste inchangée dans le cadre de ces modèles lorsqu'on utilise des résidus basés sur des estimateurs convergents et asymptotiquement normaux.

Ducharme et Lafaye de Micheaux (2004) ont construit le premier test d'ajustement spécifiquement destiné aux résidus dans les modèles ARMA avec une moyenne présumée connue. Ils utilisent l'approche des tests lisses de Neyman (1937) et Ledwina (1994). Les estimateurs des paramètres inconnus sont obtenus par maximum de vraisemblance conditionnelle dans les développements théoriques. Des résultats intéressants concernant ce type de test statistique sont disponibles dans d'autres domaines, notamment dans l'étude des données axiales et directionnelles (voir Boulerice et Ducharme (1997)), l'étude des courbes de croissance ainsi que certains modèles monotones non linéaires (voir Ducharme et Fontez (2004)), l'ajustement dans les prévisions multivariées (voir par exemple Ghosh (2003, chap. 4) et De Gooijer (2007)), la modélisation des valeurs extrêmes (Radouane et Crétois (2002), et aussi Falk, Guillou et Toulemonde (2008)), la détection de l'effet aléatoire dans les modèles linéaires mixtes (voir Thas (2009)), le problème d'homogénéité de deux échantillons dans les modèles de survie (voir Kraus (2009)) et l'ajustement à une distribution Gamma (voir De Boeck et al. (2011)).

Plus récemment, certaines généralisations dans un cadre des séries chronologiques ont été faites pour les modèles autorégressifs conditionnellement hétéroscédastiques (ARCH), et leur extension aux modèles autorégressifs conditionnellement hétéroscédastiques généralisés (GARCH). Voir par exemple Starwinski (2009).

STRUCTURE DE LA THÈSE

L'objectif principal de la recherche présentée ici est de fournir des outils statistiques qui permettent la vérification de l'adéquation d'une série chronologique à la distribution sous-jacente, particulièrement la loi normale. Nous nous sommes limités à la classe des séries chronologiques linéaires de type autorégressifs à moyenne mobile. Les objectifs spécifiques sont abordés dans différents projets que nous détaillons maintenant.

Le premier projet qui fera l'objet du chapitre 3, se situe dans le contexte des tests de normalité dans les modèles ARMA utilisant les tests lisses de Neyman. Nous avons travaillé à généraliser les travaux de Ducharme et Lafaye de Micheaux (2004) dans le cas où la moyenne est inconnue. Plusieurs méthodes d'estimation exactes et conditionnelles ont été étudiées. Des efforts ont été consacrés à l'étude d'une méthode d'estimation de type moindres carrés décrite en détail dans Brockwell et Davis (1991, section 8.11). Ainsi, le premier projet étend dans un premier temps le travail existant dans le cas univarié pour les séries chronologiques ARMA, initié par Ducharme et Lafaye de Micheaux (2004). Nous utiliserons comme famille d'emboîtement une famille exponentielle définie par les polynômes de Legendre. Il est possible de construire un test statistique avec une distribution asymptotique de loi khi-deux. Cette statistique est exprimée comme une somme de composantes indépendantes, chacune détectant un certain type d'écart à la normalité. Il est à noter que dans Ducharme et Lafaye de Micheaux (2004), la moyenne du processus ARMA était présumée connue. Nous traitons dans ce projet le cas plus réaliste où la moyenne est inconnue. La généralisation au cas où la moyenne est inconnue n'est pas triviale. En plus de nécessiter un traitement mathématique délicat, la loi de la statistique est affectée lorsque la moyenne est inconnue et doit donc être estimée. En effet, on va noter la présence d'un terme supplémentaire dans la matrice des variances et des covariances clairement associé à l'effet de l'estimation de la moyenne inconnue. Nous étudions également dans ce projet un critère de sélection de la dimension optimale basé sur les paramètres de la statistique de test. Finalement, une preuve rigoureuse de l'inversibilité de la matrice des variances et des covariances de la statistique de test, basée sur les propriétés des familles orthonormales, est étudiée en détail.

Le second projet présenté au chapitre 4 consiste à étendre les résultats du premier projet au cas vectoriel. Spécifiquement, nous proposons un test d'ajustement pour les modèles autorégressifs à moyenne mobile avec une paramétrisation structurée. Contrairement au cas univarié, le nombre de paramètres devient très

élevé dans le cas multivarié si la dimension des données est grande ou si on choisit des ordres autorégressifs et moyennes mobiles grands. La paramétrisation structurée permet de réduire le nombre élevé de paramètres dans ces modèles ou encore de tenir compte de certaines contraintes particulières. Il est intéressant de noter que la paramétrisation structurée trouve tout son sens dans un contexte multivarié. Nous utiliserons comme famille d’emboîtement une famille exponentielle impliquant des produits de polynôme d’une variable réelle (voir Rayner et Best (1989)). Le test que nous proposons s’applique à une famille quelconque de fonctions orthogonales, et nous illustrons les résultats obtenus dans les cas particuliers des polynômes de Legendre et d’Hermite. Pour les polynômes d’Hermite, nous montrons que le test obtenu est invariant aux transformations affines. Ce travail constitue ainsi une généralisation du travail de Jarque et Bera (1987) pour le test de normalité dans les données indépendantes ou encore celui de Lütkepohl (2005, p. 174) dans les modèles autorégressifs vectoriels. Par contre, dans le cas des polynômes de Legendre, une étude numérique nous révèle que le test obtenu n’est pas invariant aux transformations affines bien que possédant d’excellentes propriétés asymptotiques et à distance finie. Dans ce projet, nous supposons que la moyenne de la série est inconnue et estimée. Ceci permet une suite cohérente du premier projet et un traitement unifié.

Dans chacun des projets, une étude de simulation est réalisée afin d’étudier les propriétés asymptotiques trouvées. Dans ces études, nous analysons le niveau et la puissance de chacun des tests, en faisant notamment des comparaisons avec les tests compétiteurs. Nous illustrons également dans chacun des cas une application aux données réelles. Pour une compréhension facile de la terminologie utilisée dans ces projets, nous présentons en premier lieu certaines notions et concepts. Dans le chapitre 1, nous présentons les outils les plus fondamentaux dans l’analyse des séries chronologiques univariées et vectorielles. L’approche des tests lisses de Neyman basée sur les polynômes orthogonaux est discutée dans le chapitre 2, ainsi que quelques notions d’algèbre linéaire et d’analyse fonctionnelle. Les principales remarques et perspectives sont présentées en guise de conclusion générale.

Chapitre 1

INTRODUCTION AUX MODÈLES DE SÉRIES CHRONOLOGIQUES

Exposé simplement, une série chronologique est une suite d'observations indexées selon l'ordre temporel, chacune des observations étant supposée être la réalisation d'une certaine variable aléatoire observée à des instants équidistants. Le problème fondamental consiste à formuler des modèles parcimonieux aux données. Les applications sont nombreuses, notamment en économie, finance, hydrologie, etc. La figure 1.1 représente les déviations de la température moyenne du globe terrestre pour la période 1880 à 2009, par rapport à la température moyenne pour la période 1951 à 1980 (Shumway et Stoffer (2011, p. 5)). Les données sont exprimées en Celsius. Ce genre de séries chronologiques a été étudié afin de vérifier l'hypothèse de réchauffement climatique. La pente positive qui peut être dégagée de ce graphique est ce que l'on appelle une composante de tendance.

L'étude statistique des séries chronologiques tient compte que les données observées sont issues de variables aléatoires qui représentent des formes plus ou moins complexes de dépendance. Ainsi, de nombreux modèles expriment une variable dépendante expliquée par son propre passé. À cela peuvent s'ajouter des variables explicatives, souvent appelées variables exogènes dans la littérature portant sur les séries chronologiques. L'objectif de ce chapitre est de présenter brièvement les propriétés et les terminologies essentielles. Des approfondissements se trouvent par exemple dans le livre de Gouriéroux et Monfort (1995). Dans un premier temps, nous exposons les définitions et les concepts les plus utiles dans le contexte univarié et nous présentons la généralisation au cas vectoriel, c'est-à-dire lorsque les observations sont en fait des vecteurs aléatoires. Par la suite, la classe des modèles autorégressifs à moyenne mobile (ARMA) est introduite. Nous discutons également des méthodes générales d'estimation dans cette classe de modèles. La dernière section de ce chapitre discute des modèles ARMA vectoriels