

# 1 Curriculum Vitæ

Jean-Armel BRA

🏠 128 B rue de Belfort  
25000 Besançon, France

☎ +33 7 53 51 17 94

✉ [kja.bra@umlp.fr](mailto:kja.bra@umlp.fr)

✉ [brajeanarmel7@gmail.com](mailto:brajeanarmel7@gmail.com)

🌐 <https://bra.pages.math.cnrs.fr>

🏠 **Adresse professionnelle**

Université Marie et Louis Pasteur  
Laboratoire de Mathématiques de Besançon  
16 route de Gray, 25030 Besançon, France

👤 Célibataire sans enfants

## Situation actuelle

---

**Docteur** en mathématiques appliquées option statistique au Laboratoire de Mathématiques de Besançon (LmB), Université Marie et Louis Pasteur (UMLP), depuis le 04 décembre 2025.

**Attaché Temporaire d'Enseignement et de Recherche (ATER)** à temps complet, Université Marie et Louis Pasteur (UMLP) pour l'année universitaire 2025–2026.

## Fonctions et expérience professionnelle

---

**Sept. 2025 – Août. 2026** : ATER, UMLP, Besançon.

**Oct. 2022 – Oct. 2025** : Contrat doctoral avec avenant d'enseignement (64h/année).

**Avr. 2022 – Sept. 2022** : Stage de recherche à l'Institut Élie Cartan de Lorraine, Nancy.  
Thème : Apprentissage de noyaux de fragmentation par les réseaux de neurones.

**Fév. 2022 – Avr. 2022** : Travail d'étude et de recherche à l'Institut Jean-Lamour, Nancy.  
Thème : Initiation aux méthodes semi-lagrangiennes.

**Avr. 2021 – Mai. 2021** : Travail d'étude et de recherche à l'IECL, Nancy.  
Thème : Détection de ruptures dans les courbes d'incidences de la COVID-19.

**Avr. 2018 – Sept. 2018** : Stage d'étude et de recherche en analyse harmonique abstraite, Université Félix Houphouët Boigny, Abidjan (Côte d'Ivoire).  
Thème : Représentation des algèbres hom-Lie.

## Formation et diplômes

---

**Oct. 2022 – Déc. 2025** : Doctorat en mathématiques appliquées en statistique, Université Marie et Louis Pasteur, Besançon (France).

**Titre** : Modèles de séries temporelles modulés par une chaîne de Markov cachée avec innovations dépendantes.

COMPOSITION DU JURY

### 🎓 Rapporteurs

M. Jean-Marc BARDET, Professeur, Université Paris Panthéon-Sorbonne

Mme Delphine BLANKE, Professeure, Université Avignon

### **Examineurs**

M. Hervé CARDOT, Professeur, Université Bourgogne Europe (Président du jury)  
M. Christian FRANCO, Professeur, Université de Lille et CREST-ENSAE  
Mme. Anne GEGOUT-PETIT, Professeure, Université de Lorraine et IECL  
M. Jean-Michel ZAKOIAN, Professeur, Université de Lille et CREST-ENSAE

### **Invitée**

Mme. Camelia GOGA, Professeure, Université Marie et Louis Pasteur

### **Directeurs de thèse**

M. Yacouba BOUBACAR MAÏNASSARA, Professeur, Université Polytechnique Hauts-de-France et INSA Hauts-de-France  
M. Landy RABEHASAINA, Maître de conférences, Université Marie et Louis Pasteur

**Sept. 2021 – Sept. 2022** : Master 2 Ingénierie Mathématique pour la Science des Données, Université de Lorraine, Nancy, France. (Mention : Bien).

**Thématiques étudiées** : Apprentissage supervisé et non supervisé, clustering, intelligence artificielle, fouille de données, extraction de connaissances, gestion de bases de données (SQL), visualisation, modélisation stochastique, simulation, éléments finis et décomposition de domaines.

**Encadrants de stage** : Mme. Madalina Deaconu et M. Antoine Lejay.

**Sept. 2020 – Mai. 2021** : Master 1 Ingénierie Mathématique pour la Science des Données, Université de Lorraine, Nancy, France (Mention : Bien).

**Thématiques étudiées** : Chaînes de Markov, martingales, optimisation, calcul scientifique, statistiques avancées, algorithmes, algèbre linéaire appliquée, calcul différentiel, analyse.

**Oct. 2018 – Mai. 2020** : Élève-professeur à l'École normale supérieure (ENS) d'Abidjan, formation à l'enseignement des mathématiques dans le secondaire.

**Formation pédagogique et didactique** : préparation de cours, conduite de classe, stages pratiques d'enseignement.

**Sept. 2016 – Sept. 2018** : Master 1 et Master 2 de Mathématiques Fondamentales, Université Félix Houphouët Boigny, Abidjan, Côte d'Ivoire. (Mention : Bien).

**Thématiques étudiées** : Homotopie et cohomologie, analyse harmonique, opérateurs différentiels et algèbre de Weyl, intégrales singulières, analyse fonctionnelle, analyse complexe, géométrie différentielle, algèbre commutative et compléments de topologie.

## **2 Activités d'enseignement**

---

Enseigner est pour moi une véritable passion. J'ai eu l'opportunité de transmettre mes connaissances en mathématiques à des étudiants de niveaux variés, de la Licence 1 au Master 2, dans des groupes allant d'une quarantaine d'étudiants en Licence à une douzaine ou une vingtaine en Master. Au-delà des heures d'enseignement, j'assume également diverses responsabilités pédagogiques : préparation et actualisation des supports de cours et de travaux dirigés, élaboration des sujets d'examen et de contrôle continu, participation aux jurys et aux délibérations, encadrement de projets tutorés et de stages, ainsi qu'accompagnement individualisé des étudiants. Cette expérience m'amène à adapter constamment mes méthodes pédagogiques aux profils et aux besoins des étudiants, afin de les accompagner dans leur progression et de renforcer leur confiance dans leur pratique des mathématiques.

Enseignement	Volume	Contenus principaux
<b>Licence</b>		
L2 SVT : CM / TP Statistiques descriptives	10 h	Statistiques univariées et bivariées, TP en <code>python</code> .
L1 AES : TD Statistiques descriptives	18 h	Moyenne, médiane, variance, écart-type, quantiles, graphiques.
L1 SFA aménagés : TD Outils mathématiques 2	22 h	Outils de base en algèbre et analyse : trigonométrie, intégration, nombres complexes, équations différentielles et fonctions de plusieurs variables.
L1 LAS SFA Physique Chimie : TD Outils mathématiques 1	54 h	Outils de base en algèbre et analyse : trigonométrie, intégration, nombres complexes, équations différentielles et fonctions de plusieurs variables.
L1 SVT : TD Outils mathématiques 1	20 h	Limites et continuité, dérivation, étude de fonctions, intégration, primitives.
L1 SVT : TD Outils mathématiques 2	20 h	Primitives, intégrales, équations différentielles linéaires, modélisation.
L2 Mathématiques : Encadrement	2 h	Aide à la rédaction, étude bibliographique, présentation.
L3 Mathématiques : Encadrement	4 h	Aide à la rédaction, étude bibliographique, présentation.
Liaison Lycée – Licence : Consolidation des acquis	15 h	Logique et raisonnement, calcul algébrique, étude de fonctions, équations différentielles linéaires, méthodes.
<b>Master</b>		
M1 Modélisation statistique : TP Python	6 h	<code>numpy</code> , <code>matplotlib</code> , POO en Python, générateurs, manipulation de fichiers <code>csv</code> , approximation numérique de $\pi$ , simulation Monte Carlo, visualisation de données, loi de Benford.
M2 Modélisation statistique : TD Apprentissage statistique	9 h	Régression, classification, évaluation, validation croisée, régularisation.
M2 Modélisation statistique : TP Chaînes de Markov	3 h	Simulation de chaînes de Markov cachées, décodage des états cachés par l'algorithme de Viterbi, applications, introduction aux processus de branchement de Galton–Watson.
M2 Modélisation statistique : CM/TP Python avancé	9 h	Rappels avancés de Python, calcul scientifique avec <code>numpy</code> , visualisation avec <code>matplotlib</code> et <code>seaborn</code> , manipulation et restructuration de données avec <code>pandas</code> , introduction à <code>scikit-learn</code> et POO en Python.

Sept. 2024 – Juin. 2025

Moniteur, UMLP

Enseignement	Volume	Contenus principaux
<b>Licence</b>		
Liaison Lycée–Licence : Remise à niveau	30 h	Algèbre, fonctions, logique et raisonnement, méthodes.
L3 / M1 : Encadrement de projets	6 h	Aide à la rédaction, étude bibliographique, présentation.
<b>Master</b>		
M2 Modélisation statistique : TD Apprentissage statistique	6 h	Régression, classification, évaluation, validation croisée, régularisation.
M2 Modélisation statistique : TP Chaînes de Markov	6 h	Simulation de chaînes de Markov cachées, décodage des états cachés par l'algorithme de Viterbi, applications, introduction aux processus de branchement de Galton–Watson.
M1 Modélisation statistique : TP Python avancé	6 h	<code>numpy</code> , <code>matplotlib</code> , classes, générateurs, manipulation de fichiers <code>csv</code> , approximation numérique de $\pi$ , simulation Monte Carlo, visualisation de données, loi de Benford.

Sept. 2023 – Juin. 2024

Moniteur, UMLP

Enseignement	Volume	Contenus principaux
<b>Licence</b>		
L1 SVT : Cours et TD Outils mathématiques 1	40 h	Fonctions, limites, continuité, dérivation, applications.
L1 SVT : Cours et TD Outils mathématiques 2	10 h	Équations différentielles linéaires, primitives, intégrales, applications.
<b>Master</b>		
M2 Modélisation statistique : TP Chaînes de Markov	6 h	Simulation de chaînes de Markov cachées, décodage des états cachés par l'algorithme de Viterbi, applications, introduction aux processus de branchement de Galton–Watson.
M1 Modélisation statistique : TP Python avancé	6 h	<code>numpy</code> , <code>matplotlib</code> , classes, générateurs, manipulation de fichiers <code>csv</code> , approximation numérique de $\pi$ , simulation Monte Carlo, visualisation de données, loi de Benford.

Enseignement	Volume	Contenus principaux
<b>Licence</b>		
L1 SVT : Cours et TD Outils mathématiques 1	40 h	Limites et continuité, dérivation, étude de fonctions, intégration, primitives.
L1 SVT : Cours et TD Outils mathématiques 2	20 h	Primitives, intégrales, équations différentielles linéaires, modélisation.

### 3 Activités scientifiques

#### 3.1 Motivations et description

Lorsque l'on ajuste des modèles ARMA (AutoRegressive Moving Average) sur des sous-périodes d'une longue série temporelle (financière, par exemple), on observe souvent des variations significatives dans l'estimation des paramètres. Ces variations traduisent des changements dans la dynamique sous-jacente des données, pouvant être associés à des cycles économiques, tels que les phases de croissance et de récession. Une approche naturelle pour capturer ces changements est de recourir aux modèles à changements de régimes markoviens, où les transitions entre régimes sont décrites par une chaîne de Markov cachée. Cette chaîne est dite cachée car les états latents ne sont pas directement observés, mais influencent indirectement les paramètres du modèle, générant ainsi des dynamiques (économiques) distinctes. Ces modèles peuvent être considérés comme des mélanges markoviens de systèmes dynamiques, appartenant à la classe des modèles à changements de régimes markoviens (voir par exemple Francq et Rousignol (1997), Hamilton (1990)). Au-delà de la présence de régimes dynamiques, un enjeu majeur est la prise en compte de la dépendance entre les observations. Par exemple les mesures moyennes mensuelles du taux de  $CO_2$  présentent également une dépendance liée aux variations saisonnières, avec des fluctuations autour d'une tendance globale croissante. Les taux de pollution aux particules fines (PM2.5) dans les grandes villes montrent une dépendance à court terme, notamment en fonction des conditions météorologiques et des activités humaines. Ces exemples montrent que même en l'absence de corrélation forte, les observations successives conservent une certaine forme de dépendance ou de structure résiduelle, souvent négligée par les modèles classiques. Pour de tels modèles, il faut noter que les modèles autoregressifs standard avec erreurs indépendantes identiquement distribuées (i.i.d) sont mal adaptés.

La problématique de ma thèse porte sur l'analyse des modèles autorégressifs faibles modulés par une chaîne de Markov cachée ou en anglais AutoRegressive Hidden Markov Chain (ARHMC pour abrégé).

On dit qu'un processus  $(X_t)_{t \in \mathbb{Z}}$  admet une représentation ARHMC( $p$ ) faible modulé par une chaîne de Markov cachée  $(\delta_t)_{t \in \mathbb{Z}}$  si, pour tout  $t \in \mathbb{Z}$ , on a :

$$X_t = \sum_{i=1}^p a_i^0(\delta_t) X_{t-i} + f^0(\delta_t) \eta_t, \quad (1)$$

où  $(\eta_t)_{t \in \mathbb{Z}}$  est un processus stationnaire satisfaisant :

$$\mathbb{E}(\eta_t) = 0, \quad \mathbb{E}(\eta_t \eta_{t'}) = \sigma_0^2 \mathbf{1}_{\{t=t'\}}, \quad (2)$$

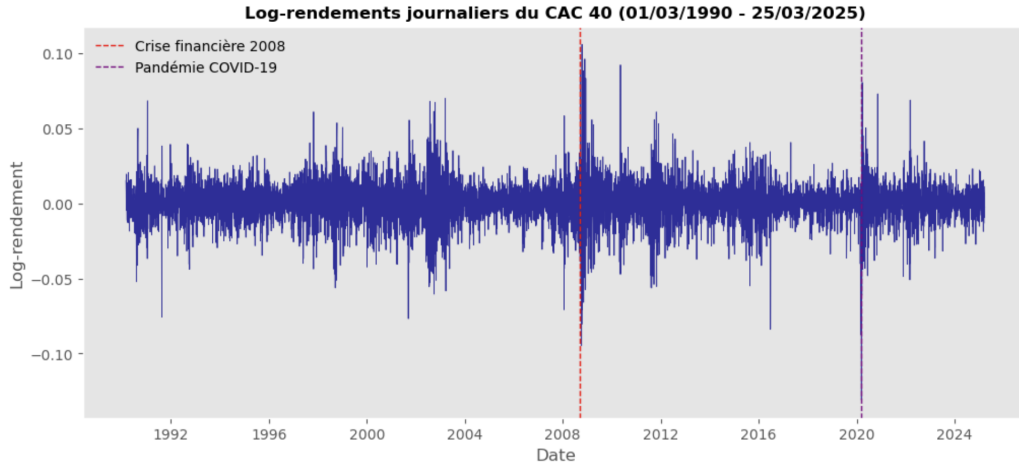
et  $(\delta_t)_{t \in \mathbb{Z}}$  est une chaîne de Markov cachée, supposée stationnaire, ergodique et irréductible, à valeurs dans  $\mathcal{E} = \{1, \dots, K\}$  de distribution stationnaire  $\pi$  et de probabilité de transition  $p^0(k, k') = \mathbb{P}(\delta_t = k' | \delta_{t-1} = k)$ ,  $f : \mathcal{E} \rightarrow \mathbb{R}$  est une fonction associant un état caché à un paramètre multiplicatif et  $(a^0(s), s \in \mathcal{E})$  est une famille de coefficients autorégressifs dépendant de l'état caché. Le vecteur de paramètres d'intérêt est défini par :

$$\theta_0 := (\sigma_0^2, a_i^0(s), p^0(k, k'), f^0(s); i = 1, \dots, p, s, k = 1, \dots, K, k' = 2, \dots, K)$$

qui appartient à l'espace des paramètres

$$\Theta := \left\{ (\sigma^2, a_i(s), p(k, k'), f(s); i = 1, \dots, p, s, k = 1, \dots, K, k' = 2, \dots, K) \right. \\ \left. \in \mathbb{R}_+ \times \mathbb{R}^{pK} \times [0, 1]^{K(K-1)} \times \mathbb{R}^K; P_\theta := (p(k', k)) \text{ irréductible} \right\}.$$

Le processus ainsi défini dans l'équation (2) est appelé **bruit blanc faible**. Il est important de noter qu'aucune hypothèse d'indépendance n'est faite dans la définition du **bruit blanc faible**, les variables aux différentes dates sont simplement non corrélées. Un exemple de processus pouvant être ajusté à des **bruits blancs faibles**, dont les autocorrélations sont nulles mais qui peuvent néanmoins présenter des dépendances temporelles, est celui des processus autorégressifs conditionnellement hétéroscédastiques (ARCH pour AutoRegressive Conditional Heteroskedasticity). Ainsi que leurs extensions GARCH (Generalized ARCH). Les séries générées par ces modèles GARCH sont des différences de martingales et sont utilisées pour décrire des séries financières. Dans la pratique, il existe de nombreuses situations où l'hypothèse d'erreurs i.i.d. gaussiennes, également appelée **bruit blanc fort**, n'est pas vérifiée. À d'exemple illustratif de tel processus défini en (2), nous pouvons considérer le graphique suivant de la série du rendement du CAC40, défini par  $\eta_t = \log(p_t/p_{t-1})$ , où  $p_t$  est le prix de l'action du CAC40 à la date  $t$ .



La série du rendement du CAC40 présente une trajectoire apparemment compatible avec la stationnarité au second ordre. On constate sur le graphique que la série  $(\eta_t)$  est constituée d'oscillations autour de zéro, avec des amplitudes très différentes d'une date à l'autre, mais de module moyen constant sur des sous-périodes suffisamment longues. On observe également que la série  $(\eta_t)$  présente de très faibles autocorrélations, la rendant proche d'un bruit blanc. En revanche, la série des carrés  $(\eta_t^2)$  ou la série des valeurs absolues  $(|\eta_t|)$  sont souvent fortement autocorrélées, ce qui prouve aussi qu'une dynamique peut exister au niveau du **bruit blanc faible** contrairement au bruit blanc fort. Ces deux propriétés ne sont pas incompatibles, mais elles montrent que la série du rendement est un bruit blanc qui n'est pas fort.

### 3.2 Description des travaux réalisés

Dans cette sous-section, je propose un résumé des travaux que j'ai réalisés en collaboration avec Y. Boubacar Maïnassara et L. Rabehasaina. Pour en faciliter la lecture et la référence, ces travaux seront désignés par **[P1]**, **[P2]**, **[P3]** et **[P4]** dans la section « Prépublication » suivante.

**[P1]** : Dans un premier temps, nous considérons le modèle (1) avec  $p = 1$ . Sous les hypothèses essentielles d'ergodicité et de mélange, ainsi que d'indépendance entre  $(\eta_t)$  et  $(\delta_t)$ , nous avons proposé un estimateur implicite de  $\theta_0$  par la **méthode des moments**, puis implémenté cette approche (en **Python**) via un algorithme de quasi-Newton.

Sous ces hypothèses, les paramètres du modèle (1) satisfont, pour tout  $k \in \mathbb{N}^*$ ,

$$c_{k,0}(\theta_0) := \mathbb{E}[X_k X_0] = \sigma_0^2 \mathbf{1}^\top (A_{\theta_0} P_{\theta_0}^\top)^k A_{\theta_0}^2 P_{\theta_0}^\top (I - A_{\theta_0}^2 P_{\theta_0}^\top)^{-1} \boldsymbol{\pi}_{f_0^2},$$

où  $\mathbf{1}^\top := (1, \dots, 1)$ ,  $\boldsymbol{\pi}_{f_0^2} := \{f_0^2(1)\pi(1), \dots, f_0^2(K)\pi(K)\}^\top$ ,  $A_{\theta_0} := \text{diag}(a^0(s), s \in \mathcal{E})$  et  $P_{\theta_0} = (p_{i,j}^0)$ . En supposant que nous disposons d'observations  $X_1, \dots, X_n$ ;  $n \in \mathbb{N}$ , une *fonction d'estimation* du paramètre  $\theta_0$  est la fonction  $F^{N,n}(\cdot)$  définie par

$$\mathcal{F}^{N,n}(\theta) = [\nabla F^{N,n}(\theta)]^\top F^{N,n}(\theta), \quad \theta \in \Theta$$

où

$$F^{N,n} := (\hat{c}_{0,0} - c_{0,0}(\theta), \dots, \hat{c}_{N-1,0} - c_{N-1,0}(\theta))^\top,$$

avec  $\hat{c}_{k,0}$  est un estimateur empirique de  $c_{k,0}(\theta_0)$  et  $N$  est un entier naturel représentant le nombre de moments croisés considérés ( $N \geq K^2 + K + 1$ , le nombre de paramètres à estimer).

Un estimateur  $\hat{\theta}_n$  de  $\theta_0$  est alors obtenu en annulant la fonction  $\mathcal{F}^{N,n}(\cdot)$ . Nous avons ensuite étudié les **propriétés asymptotiques** de l'estimateur proposé. Sous les mêmes hypothèses (existence des  $c_{k,0}$  et l'hypothèse que  $J := (\nabla c_{0,0}(\theta_0), \dots, \nabla c_{N-1,0}(\theta_0))^\top$  est de plein rang), l'estimateur  $\hat{\theta}_n$  converge vers  $\theta_0$  et vérifie le **théorème central limite** suivant

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \Omega := \mathcal{G}^{-1} J^\top I J^\top \mathcal{G}^{-1}),$$

où les matrices  $\mathcal{G}$  et  $I$  sont définies comme suit

$$\mathcal{G} := J^\top J \quad \text{et} \quad I := \sum_{k=-\infty}^{\infty} \text{Cov}(Y_t(\theta_0), Y_{t-k}(\theta_0)) \quad \text{avec} \quad Y_t := X_t(X_t, X_{t+1}, \dots, X_{t+N})^\top.$$

Nous nous sommes aussi intéressés à l'**estimation de la matrice de variance asymptotique**  $\Omega$ . Pour cela, nous avons estimé la matrice  $J$  par **plug-in** en remplaçant  $\theta_0$  par  $\hat{\theta}_n$ . L'estimation de la matrice  $I$  est un peu plus délicate. Nous avons choisi de nous focaliser sur l'approche basée sur l'estimation par la **densité spectrale**. Cette méthode consiste à interpréter  $(2\pi)^{-1}I$  comme la densité spectrale du processus  $\{\mathcal{Y}_t := Y_t - \mathbb{E}(Y_t), t \in \mathbb{Z}\}$ .

L'estimateur proposé est donné par :

$$\hat{I}^{SP} := \hat{\varphi}(1)^{-1} \hat{\Sigma}_{\hat{u}_r} \hat{\varphi}'(1)^{-1} \rightarrow I$$

en probabilité quand  $r = r(n) \rightarrow \infty$  et  $r^3/n \rightarrow 0$  quand  $n \rightarrow \infty$ , avec  $\hat{\varphi}(z) = I_N + \sum_{i=1}^r \hat{\varphi}_{r,i} z^i$  où  $\hat{\varphi}_{r,1}, \dots, \hat{\varphi}_{r,r}$  les coefficients de la régression des moindres carrés de  $\hat{\mathcal{Y}}_t$  sur  $\hat{\mathcal{Y}}_{t-1}, \dots, \hat{\mathcal{Y}}_{t-r}$  et  $\hat{\Sigma}_{\hat{u}_r}$  la variance empirique des résidus de cette régression. Enfin, nous avons appliqués nos résultats sur des données météorologiques horaires pour corroborer nos travaux théoriques.

**[P2]** : Nous avons étudié le modèle (1) pour un ordre  $p$  quelconque. Nous avons estimé le paramètre d'intérêt et démontré les propriétés asymptotiques de l'estimateur obtenu. En outre, nous avons déterminé la **matrice de variance asymptotique** et généralisé la procédure d'estimation à un processus  $(X_t)_{t \in \mathbb{Z}}$  potentiellement non centré. L'un des principaux défis de cette étude a été de reformuler le modèle (1) sous la forme d'un processus multivarié d'ordre 1. Cette reformulation était essentielle pour permettre le calcul des moments croisés. Pour ce faire, il a été nécessaire d'étendre le Lemme 1 de Francq et Gautier (2004) au cas multidimensionnel. Ainsi, l'ensemble de notre approche a consisté à généraliser le cadre d'étude initial en exprimant le modèle sous forme de vecteur autorégressif. Cette formulation a permis de traiter les propriétés asymptotiques dans un contexte plus vaste et plus rigoureux. Une attention particulière a également été accordée à l'application de nos résultats à des données réelles dans ce cadre.

**[P3]** : Nous avons également étudié le cas particulier  $p = 0$ , où dans ce cas le processus  $(X_t)_{t \in \mathbb{Z}}$  est défini par :

$$X_t = f^0(\delta_t)\eta_t.$$

À noter que la procédure utilisée dans la section précédente ne convient pas au cas ARHMC(0). En effet, les moments croisés sont nuls dans cette configuration en raison de l'indépendance entre  $(\delta_t)_{t \in \mathbb{Z}}$  et  $(\eta_t)_{t \in \mathbb{Z}}$ , ainsi que du fait que  $(\eta_t)_{t \in \mathbb{Z}}$  est centré. Par conséquent, nous avons dû explorer les moments croisés d'ordre 4 de  $(X_t)_{t \in \mathbb{Z}}$  afin de contourner cette difficulté. Après la phase d'estimation du processus ARHMC(0), la prochaine étape importante dans la modélisation consiste à vérifier si le modèle estimé est compatible avec les données. Cette étape d'adéquation permet de valider ou d'invalider le choix de  $K$  du nombre de régimes du modèle. Ce choix est important pour la précision des prévisions et pour une bonne interprétation du modèle. La sélection d'un ordre  $K$  trop grand a pour effet d'introduire des termes qui ne sont pas forcément pertinents dans le modèle et aussi d'engendrer des difficultés statistiques comme, par exemple un trop grand nombre de paramètres à estimer, ce qui est susceptible d'engendrer une perte de précision de l'estimation des paramètres. Le praticien peut aussi choisir un ordre  $K$  trop petit qui entraîne la perte d'une information qui peut être détectée par une corrélation des résidus ou encore une estimation non convergente des paramètres. Nous avons développé des tests adaptés aux erreurs faibles. Ces tests ont pour objet de vérifier que les résidus des modèles estimés sont bien des estimations de bruits blancs. Une invalidation de cette condition implique que le modèle est mal spécifié. Plus particulièrement, nous nous intéressons aux tests portmanteau, aussi appelés tests d'autocorrélation. Dans un premier temps, dans **[P3]**, nous avons étudié la distribution asymptotique jointe de l'estimateur des moments et des autocovariances empiriques de  $(\eta_t)_{t \in \mathbb{Z}}$ . Ceci nous permet ensuite d'obtenir les distributions asymptotiques des autocovariances et autocorrélations résiduelles par la delta-méthode. Nous avons prouvé que ces autocorrélations résiduelles sont normalement distribuées avec une matrice de covariance différente du cas iid. Enfin, nous en avons déduit le comportement asymptotique des statistiques portmanteau. Dans le cas général, nous avons obtenu la loi exacte des tests portmanteau et nous avons également montré que cette distribution asymptotique est celle d'une somme pondérée de chi-deux. Nous en avons déduit des tests portmanteau modifiés pour tester l'adéquation du modèle ARHMC(0) faible. Comme conséquence des résultats obtenus, il a fallu adapter les bornes des autocorrélogrammes en présence de dépendance. Après l'étape de **validation**, une phase cruciale est celle de la **prévision**, qui consiste à estimer les valeurs futures de la série à partir du modèle ajusté précédemment. Cette étape ne se limite pas à une simple prédiction ponctuelle, mais inclut également l'évaluation de l'incertitude via des intervalles de confiance, garantissant ainsi la robustesse des prévisions. Cependant, dans notre contexte, la tâche s'est avérée particulièrement complexe en raison de l'absence d'hypothèse de loi sur le processus  $(\eta_t)_{t \in \mathbb{Z}}$ . En effet, les approches standard utilisées lorsque  $(\eta_t)_{t \in \mathbb{Z}}$  est i.i.d. ne pouvaient être appliquées ici. La principale difficulté résidait donc dans la détermination de la densité du processus  $(X_t)_{t \in \mathbb{Z}}$ . Pour

surmonter cet obstacle, nous avons adopté une approche innovante consistant à approximer la densité jointe de  $(\eta_t)_{t \in \mathbb{Z}}$  à l'aide des polynômes de **Legendre**. Cette technique s'est révélée non seulement efficace pour l'estimation de la densité, mais elle nous a également permis, par la même occasion, de décoder la trajectoire de la chaîne latente  $(\delta_t)_{t \in \mathbb{Z}}$ . Cette double avancée nous a permis de contourner les limitations imposées par l'absence d'hypothèse de loi, tout en offrant une meilleure interprétation des dynamiques sous-jacentes.

[P4] : Un autre volet de notre travail porte sur les modèles fGARCH (Functional GARCH models). Nous y introduisons un fGARCH(1,1) modulé par une chaîne de Markov cachée en espace de Hilbert, en établissons la bonne définition (stationnarité stricte) ainsi que l'existence de moments d'ordre deux sous des conditions simples, et récrivons la dynamique sous forme bloc-opérateur. La projection sur une base finie fournit alors une représentation matricielle exploitable. À partir de celle-ci, nous dérivons des relations fermées entre paramètres et moments croisés des courbes au carré, et construisons une fonction d'estimation en appariant moments théoriques et empiriques.

### 3.3 Publication et Prépublication

[P1] Y. Boubacar Maïnassara, A. Bra et L. Rabehasaina (2025). *Estimating Weak Markov-Switching AR(1) Models* (publié à Electronic Journal of Statistics (EJS)). arXiv:2503.03316

[P2] Y. Boubacar Maïnassara, A. Bra et L. Rabehasaina, (2025). *Estimating Weak Markov-Switching AR(p) Models* (en révision à Statistical Papers) .

[P3] Y. Boubacar Maïnassara, A. Bra et L. Rabehasaina (2025). *Modelling on weak white noise driven by a hidden Markov chain* (en cours de finalisation).

[P4] Y. Boubacar Maïnassara, A. Bra et L. Rabehasaina (2025). *Functional GARCH models driven by a hidden Markov chain* (en cours de finalisation)

### 3.4 Perspectives de recherche

Mes travaux de thèse s'inscrivent dans le domaine de la statistique des séries temporelles, avec un intérêt particulier pour les modèles autorégressifs à changements de régime markoviens. Plus précisément, je me suis intéressé à des modèles ARHMC (AutoRegressive Hidden Markov Chain) faibles, dans lesquels les innovations sont supposées non corrélées mais non nécessairement indépendantes. Ce cadre permet d'élargir de manière significative le champ d'application des modèles à changements de régime, tout en soulevant des questions théoriques et méthodologiques nouvelles en matière d'estimation, d'identifiabilité et d'inférence. Les résultats obtenus au cours de la thèse portent notamment sur l'estimation par la méthode des moments, l'étude des propriétés asymptotiques des estimateurs, le décodage des trajectoires latentes, la prévision ainsi que leur mise en oeuvre sur données simulées et réelles. Dans le prolongement direct de ces travaux, je souhaite développer un projet de recherche structuré autour de plusieurs axes complémentaires.

**Axe 1 : Modèles fGARCH à changements de régime markoviens.** Un premier axe de recherche portera sur l'analyse de modèles fGARCH (functional Generalized AutoRegressive Conditional Heteroskedasticity) dont les paramètres sont pilotés par une chaîne de Markov cachée. Ce travail, amorcé durant ma thèse, concerne plus précisément des modèles fGARCH(1,1)-HMM dans lesquels la dynamique de volatilité est décrite en dimension fonctionnelle, sous l'hypothèse d'innovations i.i.d.. Ce cadre combine ainsi la structure des modèles cachés à changements de régime avec une modélisation fonctionnelle explicite de la variance conditionnelle. Sur le plan méthodologique, l'objectif sera d'exploiter les expressions fermées des moments croisés

des projections afin de construire des procédures d'estimation par la méthode des moments adaptées à ces modèles. À court terme, ce travail vise à être finalisé, en particulier par l'étude des propriétés asymptotiques de l'estimateur proposé, notamment sa consistance et sa normalité asymptotique, ainsi que par l'estimation de la matrice de variance asymptotique. Cet axe permettra de prolonger mes travaux vers le cadre des séries temporelles fonctionnelles, tout en conservant une forte cohérence avec les problématiques de dépendance, d'inférence et de changements de régime développées dans la thèse.

**Axe 2 : Extension des modèles ARHMC faibles à des cadres plus généraux.** Un deuxième axe de recherche consistera à étendre les méthodes développées dans la thèse à des modèles plus riches, en particulier aux modèles VARHMC (Vector Autoregressive Hidden Markov Chain) faibles et ARMAHMC (Autoregressive Moving Average Hidden Markov Chain) faibles. Dans le cas vectoriel, l'enjeu sera de formuler des conditions de moments adaptées à une structure matricielle, d'établir des conditions d'identifiabilité et de stabilité, puis de dériver les propriétés asymptotiques des estimateurs associés. Dans le cas ARMAHMC (Autoregressive Moving Average Hidden Markov Chain), l'introduction d'une composante moyenne mobile soulève des questions supplémentaires, notamment en termes d'inversibilité et d'identifiabilité. Une piste naturelle consistera à exploiter des approximations de type ARHMC( $\infty$ ) tronquées afin de construire des moments exploitables. Cet axe vise ainsi à élargir le cadre théorique d'étude des modèles faibles à changements de régime.

**Axe 3 : Diagnostic, sélection de modèle et prévision.** Un troisième axe portera sur les questions d'inférence post-estimation pour les modèles ARHMC faibles d'ordre  $p$ . Il s'agira notamment de généraliser les tests portmanteau modifiés au cas  $p \geq 1$ , en tenant compte de la dépendance résiduelle induite par la dynamique autorégressive et par la présence d'une chaîne latente. L'objectif sera d'étudier la loi limite jointe des estimateurs et des autocorrélations résiduelles, afin de construire des outils de diagnostic adaptés à ce cadre. Ces développements pourront être mobilisés pour la sélection conjointe du nombre de régimes  $K$ , de l'ordre autorégressif  $p$ , ainsi que du nombre de moments utilisés à l'estimation. Cet axe inclut également une composante prévisionnelle, fondée sur la construction de lois prédictives approchées intégrant à la fois la dynamique autorégressive et les probabilités de régime issues de procédures de filtrage et de lissage adaptées.

**Axe 4 : Estimation robuste de la variance asymptotique.** Un quatrième axe concernera l'estimation de la matrice de variance asymptotique des estimateurs dans le cadre faible. Au-delà de l'approche fondée sur la densité spectrale à fréquence nulle utilisée dans la thèse, je souhaite étudier des méthodes non paramétriques de type HAC (Heteroskedasticity and Autocorrelation Consistent), qui permettent d'estimer directement la variance de long terme sans imposer de structure paramétrique supplémentaire à la dépendance. L'objectif sera d'établir des conditions de validité théorique de ces procédures dans le cadre ARHMC faible, d'en comparer les performances à celles des approches spectrales, et d'en évaluer l'intérêt pratique pour l'inférence en échantillon fini.

**Axe 5 : Estimation par pseudo-vraisemblance et méthodes alternatives.** Enfin, un cinquième axe visera à développer des procédures d'estimation alternatives à la méthode des moments, en particulier des approches de type pseudo-maximum de vraisemblance. Dans le cadre fort, les modèles à changements de régime se prêtent naturellement à des procédures de type EM. Dans le cadre faible, la dépendance temporelle des innovations rend cette factorisation plus délicate et appelle de nouvelles stratégies d'approximation. Une piste de travail consistera à adapter les schémas de filtrage et de lissage à un cadre de pseudo-vraisemblance, éventuellement en s'appuyant sur des estimations non paramétriques de densités conditionnelles. Cet

axe permettra d'explorer de nouvelles méthodes d'inférence, complémentaires à l'approche par moments, et d'en comparer les propriétés théoriques et numériques.

### 3.5 Communications

#### ▣ Congrès et conférences

- Congrès des jeunes statisticiens : présentation de travaux, Porquerolles 31 mars 2024
- 55<sup>es</sup> Journées de Statistique : présentation de travaux, Bordeaux 27 mai 2024
- 55<sup>es</sup> Journées de Statistique : présentation de travaux, Marseille 2 juin 2025

#### 🗣 Séminaires

- Séminaire de probabilités, analyse et statistique : exposé, Clermont-Ferrand 19 mars 2026
- Séminaire de statistique : exposé, Strasbourg 6 février 2026
- Séminaire de probabilités et statistique : exposé, Besançon 25 mars 2024
- Séminaire des doctorants : présentation des travaux de thèse, Besançon 22 mars 2024

#### 💡 Journées scientifiques et journées thématiques

- Fédération Bourgogne-Franche-Comté : exposé invité, Dijon 7 novembre 2025
- Première journée des séries temporelles : exposé invité, Valenciennes 24 juin 2025
- 22<sup>e</sup> journée de l'École doctorale Carnot–Pasteur : participation, Dijon 23 juin 2024
- Journée Probabilités–Statistiques Besançon–Dijon : participation 9 juin 2023

#### ✳ Écoles d'hiver et écoles thématiques

- École d'hiver du master de statistiques de Besançon (4<sup>ème</sup> édition) : membre de la table ronde, Villers-le-Lac 31 janvier 2024
- École d'hiver du master de statistiques de Besançon (3<sup>ème</sup> édition) : présentation, Métabief 2 février 2023

## 4 Informations complémentaires

---

### 4.1 Compétences linguistiques et informatiques

- 🌐 **Langues** : Français (langue maternelle), Anglais (courant).
- 🔗 **Logiciels scientifiques** : Python, R, MATLAB, Scilab, L<sup>A</sup>T<sub>E</sub>X, GeoGebra.
- 💻 **Bureautique** : Microsoft Word, Microsoft Excel.
- 🗄 **Outils de data science** : Weka, KNIME, MySQL.

### 4.2 Formations suivies dans le cadre de l'ED Carnot Pasteur

- 🌟 Intégrité scientifique (3 h)      🇬🇧 Anglais général (25 h)
- 📊 Modélisation statistique (10 h)      ⚖ Éthique de la recherche (30 h)
- ⚙ Se former pour enseigner dans le supérieur (30 h)      🏛 Éthique de la recherche et intégrité scientifique (25 h)

### 4.3 Centres d'intérêt

✈ Voyages, ⚽ Football, 🎵 Musique, ♟ Echecs.

### 4.4 Contacts (Recherche)

**M. Yacouba BOUBACAR MAÏNAS-SARA**  
Université Polytechnique Hauts-de-France,  
INSA Hauts-de-France CERAMATHS  
Laboratoire de Mathématiques et Matériaux  
59313 Valenciennes, France  
✉ Yacouba.BoubacarMainassara@uphf.fr

**M. Jean-Marc BARDET**  
Université Paris 1 Panthéon-Sorbonne  
SAMM – Analyse et Modélisation  
Multidisciplinaire  
90 rue de Tolbiac, 75013 Paris, France  
✉ Jean-Marc.Bardet@univ-paris1.fr

**M. Landy RABEHASAINA**  
Université Marie et Louis Pasteur  
Laboratoire de Mathématiques de Besançon  
16 route de Gray  
25030 Besançon cedex, France  
✉ landy.rabehasaina@umlp.fr

**M. Antoine LEJAY**  
Université de Lorraine  
Institut Élie Cartan de Lorraine  
54506 Vandœuvre-lès-Nancy, France  
✉ antoine.lejay@univ-lorraine.fr

### 4.5 Contacts (Enseignement)

**M. Landy RABEHASAINA**  
Université Marie et Louis Pasteur  
Laboratoire de Mathématiques de Besançon  
16 route de Gray  
25030 Besançon cedex, France  
✉ landy.rabehasaina@umlp.fr

**M. Clément DOMBRY**  
Université Marie et Louis Pasteur  
Laboratoire de Mathématiques de Besançon  
16 route de Gray  
25030 Besançon cedex, France  
✉ clement.dombry@umlp.fr

**M. Julien Yves ROLLAND**  
Université Marie et Louis Pasteur  
Laboratoire de Mathématiques de Besançon  
16 route de Gray  
25030 Besançon cedex, France  
✉ julien.rolland@umlp.fr

**Mme Camelia GOGA**  
Université Marie et Louis Pasteur  
Laboratoire de Mathématiques de Besançon  
16 route de Gray  
25030 Besançon cedex, France  
✉ camelia.goga@umlp.fr